

UNIVERSIDAD PERUANA UNIÓN
FACULTAD DE INGENIERÍA Y ARQUITECTURA
Escuela Profesional Ingeniería de Sistemas



Una Institución Adventista

**Aplicación del algoritmo k-means para clasificar el perfil del turista
según su motivación que realiza Turismo Rural Comunitario en la
Región Puno**

Tesis para obtener el Título Profesional de Ingeniero de Sistemas

Por:

Yerson Pariapaza Larico

Asesor:

Ing. Eder Gutierrez Quispe

Co-asesor:

Dra. Sc. Nélide Gladys Maquera Sosa

Juliaca, Junio del 2019

DECLARACIÓN JURADA DE AUTORÍA DEL INFORME DE TESIS

Ing. Eder Gutierrez Quispe, de la Facultad de Ingeniería y Arquitectura, Escuela Profesional de Ingeniería de Sistemas, de la Universidad Peruana Unión.

DECLARO:

Que el presente informe de investigación titulado: **“APLICACIÓN DEL ALGORITMO K-MEANS PARA CLASIFICAR EL PERFIL DEL TURISTA SEGÚN SU MOTIVACIÓN QUE REALIZA TURISMO RURAL COMUNITARIO EN LA REGIÓN PUNO”** constituye la memoria que presenta el Bachiller **Yerson Pariapaza Larico** para obtener el título Profesional de ingeniero de Sistemas, cuya tesis ha sido realizada en la Universidad Peruana Unión bajo mi dirección.

Las opiniones y declaraciones en este informe son de entera responsabilidad del autor, sin comprometer a la institución.

Y estando de acuerdo, firmo la presente declaración en Juliaca, a los 21 días del mes de agosto del año 2019.



Ing. Eder Gutierrez Quispe
Asesor



035

ACTA DE SUSTENTACIÓN DE TESIS

En Puno, Juliaca, Villa Chullunquiari, a dos día(s) del mes de julio del año 2017 siendo las 12:20 horas, se reunieron en el Salón de Grados y Títulos de la Universidad Peruana Unión, Filial Juliaca, bajo la dirección del Señor Presidente del jurado: Mg. Lenin Henry Centurión Julca, el secretario: Dr. Jorge Alejandro Sánchez García y los demás miembros: Mg. Milton Omar Santillan Asching y el asesor Ing. Eder Gutierrez Quipe

con el propósito de administrar el acto académico de sustentación de la tesis titulada: Aplicación del algoritmo K-means para clasificar el perfil del turista según su motivación que realiza Turismo Rural Comunitario en la Región Puno

de el(los)/la(las) bachiller/es: a) Yerson Pariapaza Larico b) conducente a la obtención del título profesional de Ingeniero de Sistemas (Nombre del Título Profesional)

con mención en.....

El Presidente inició el acto académico de sustentación invitando al (los)/a(la)(las) candidato(a)s hacer uso del tiempo determinado para su exposición. Concluida la exposición, el Presidente invitó a los demás miembros del jurado a efectuar las preguntas, y aclaraciones pertinentes, las cuales fueron absueltas por el(los)/la(las) candidato(a)s. Luego, se produjo un receso para las deliberaciones y la emisión del dictamen del jurado.

Posteriormente, el jurado procedió a dejar constancia escrita sobre la evaluación en la presente acta, con el dictamen siguiente:

Candidato (a): Yerson Pariapaza Larico

Table with columns: CALIFICACIÓN, ESCALAS (Vigesimal, Literal, Cualitativa), Mérito. Values: Aprobado, 16, B, Bueno, Muy bueno

Table for Candidate (b) with columns: CALIFICACIÓN, ESCALAS (Vigesimal, Literal, Cualitativa), Mérito. All cells are empty.

(*) Ver parte posterior

Finalmente, el Presidente del jurado invitó al(los)/a(la)(las) candidato(a)s a ponerse de pie, para recibir la evaluación final y concluir el acto académico de sustentación procediéndose a registrar las firmas respectivas.

Signature of Presidente

Signature of Asesor

Signature of Candidato/a (a)

Signature of Miembro

Signature of Secretario

Signature of Miembro

Signature of Candidato/a (b)

Dedicatoria

Gracias a mis padres por apoyarme siempre en el proceso de escritura. Gracias a mis hermanas, que siempre me inspiran y son modelos a seguir para las personas, y a mis amigas, que siempre me dan apoyo moral.

Agradecimientos

Gracias Dios por darme la vida para lograr una de mis otras metas, a mis amados padres por su apoyo

incansable e incondicional, mis hermanas Mónica y Sonia;

Gracias a mis asesores ya toda esta gente por su motivación y su tiempo y carácter.

Este trabajo fue financiado por el CONCYTEC-FONDECYT en el marco de la convocatoria

“Proyecto Investigación Básica 2015” [número de contrato 184-2015].

Índice General

Dedicatoria	iv
Agradecimientos	v
Índice General	vi
Índice de Tablas	ix
Índice de Figuras	x
Índice de Anexos	xi
Resumen	xii
Abstract	xiii
Capítulo I. El Problema	14
1.1. Descripción de la Situación Problemática	14
1.2. Objetivos de la investigación	16
1.2.1. Objetivo general	16
1.2.2. Objetivos específicos	17
1.3. Justificación	17
1.3.1. Social	17
1.3.2. Económico	18
1.3.3. Ambiental	18
1.3.4. Legal	19

1.3.5. Tecnológico.....	19
1.4. Presunción filosófica.....	19
Capítulo II. Marco Teórico	21
2.1. Revisión de la literatura	21
2.2. Marco teórico	29
2.2.1. Turismo.	30
2.2.2. Inteligencia Artificial	32
2.2.3. Técnicas Descriptivas o No supervisadas	33
2.2.4. Variantes del Algoritmo K-means.....	36
2.2.5. CRISP-DM.....	36
2.2.6. Herramientas de Machine Learning	38
2.2.7. Limpieza de datos.....	39
Capítulo III. Materiales y Métodos	41
3.1. Descripción del lugar de ejecución.	41
3.2. Materiales e insumos.....	41
3.3. Metodología	42
3.3.1. Tipo de la Investigación	42
3.3.2. Arquitectura de solución	42
3.3.3. Encuesta	42
3.3.4. Programación Extrema(XP).....	43

3.3.5. CRISP-DM	43
3.4. Cronograma del proyecto	44
3.5. Presupuesto	44
Capítulo IV. Resultados y discusión	45
Capítulo V. Conclusiones y Recomendaciones	48
REFERENCIAS BIBLIOGRÁFICAS	49
ANEXOS	51

Índice de Tablas

Tabla1 Revisión de la literatura	21
Tabla 2 Herramientas para Machine Learning.....	38
Tabla 3 Variables	45

Índice de Figuras

Figura 1. Proceso de Metodología ágil CRISP-DM.....	37
Figura 2. Metodología para la identificación de perfil del turista.....	42
Figura 3. Variables Definidas para el Análisis.....	46
Figura 4. Grafica del punto de codo para el numero de clúster.	47
Figura 5. Clústeres definidos gráficamente.....	47

Índice de Anexos

ANEXO A. Firma de Emprendedores del Curso Para Recolección de Datos	51
ANEXO B. Anexo B. Participación de la capacitación a emprendedores y recolección de datos de los mismos para aplicar encuesta a los turistas.	52
ANEXO C. Cuaderno Jupyter de Limpieza de Datos	53
ANEXO D. Encuesta de PromPerú.....	55
ANEXO E. MAPIC.....	57

Resumen

El objetivo de este estudio es segmentar a los turistas que viajan a las comunidades rurales de la región Puno según su motivación. El desarrollo de este proyecto fue guiado por CRISP-DM, un método estándar para realizar planes de explotación minera de datos que divide el sumario en seis (6) pasos generalmente a los específicos, a la medida de las necesidades del entorno y los algoritmos de aplicación del proyecto. . . Para demostrar el gran potencial de la técnica de agrupación de k-means, esta técnica aborda la ausencia de partes naturales en los perfiles turísticos. Trabajando sobre una colección de 1122 encuestas propiedad de turistas, utilizando técnicas de k-means, es posible encontrar similitudes en los segmentos para generar la identificación del perfil. El análisis nos permitió identificar dos grupos bien definidos de empresas para identificar perfiles.

Palabras clave: Machine Learning, Segmentación de Perfil de turista, K-Means, Turismo Rural Comunitario.

Abstract

The objective of this study is to segment tourists who travel to rural communities in the Puno region according to their motivation. The development of this project was guided by CRISP-DM, a standard method for carrying out data mining plans that divides the summary into six (6) steps, generally to the specific ones, tailored to the needs of the environment and the algorithms of project application. . . To demonstrate the great potential of the k-means clustering technique, this technique addresses the absence of natural parts in tourism profiles. Working on a collection of 1122 tourist-owned surveys, using k-means techniques, it is possible to find similarities in the segments to generate profile identification. The analysis allowed us to identify two well-defined groups of companies to identify profile.

Capítulo I. El Problema

1.1. Descripción de la Situación Problemática

Los turistas son la esencia del turismo, pero son diferentes porque tienen diferentes necesidades y compran en diferentes momentos y en diferentes cantidades. Se pueden distinguir perfiles de turistas con características muy similares a pesar de que no son homogéneos, como los países de procedencia de los turistas, su comportamiento, educación, ocupación, gustos, etc., que aportan información valiosa que, debidamente correlacionada y analizada, puede segmentar el mercado, es decir, crear grupos totalmente distinguibles, centrar los esfuerzos comerciales y de marketing en estos grupos y hacerlos más rentables con el tiempo. Los sectores altamente significativos de la mundial economía en los reciente años del turismo. (Ferreira, Rial, & Varela, 2009).

Comprender las oportunidades del mercado es crucial para avanzar y desarrollar estrategias de marketing porque el turismo es una fuente fundamental de entradas para los fines que se benefician de él. (Pereira, Anjos, & Añaña, 2017). Además de los modelos teóricos relacionados con el turismo, el primordial propósito de los estudiosos es desplegar y adaptar a las necesidades existentes de la industria turística métodos que puedan proporcionar la información de mercado necesaria (Ferreira, Rial, & Varela, 2010).

La principal organización internacional que juega un papel crucial e importante en el desarrollo de las industrias turísticas responsables, sostenibles y accesible es la Organización para el Turismo Mundial (OMT), una delegación técnica de ONU. Sirve como un foro mundial para la discusión de temas relacionados con la política del sector turístico y un germen útil de indagación turística. contiene 158 países, 6 órganos partidarios y más de 500 órganos afiliados. (Guterres, 2015).

El turismo en el Perú está en aumento año tras año, el cual genera un ingreso económico muy importante para el país, así también para las personas, distritos, municipios y regiones.

Según (Espinola, 2017) sostiene: “Entre 2011 y 2015, el PBI turístico aumentó a un ritmo superior al PBI del país. (Perú)”.

“El 2015 el Producto Bruto Interno del turismo reconoció una aportación de 3,9% conforme al PBI del país y el Empleo del Turismo Directo e Indirecto alcanzó 1,3 millones en el Perú” (Espinola, 2017).

El turista del área rural comunitario es desemejante al turismo tradicional, el cual en estos últimos años está teniendo bastante aceptación por parte de los turistas el cual también su crecimiento y difusión está en aumento según la OMT.

En el Perú, “turismo rural” se refiere a toda actividad turística realizada de manera planificada y sostenible en un área rural, basada en la participación comunitaria, organizada en beneficio de la sociedad, y donde el componente principal del producto final es la cultura rural. (mincetur, n.d.).

“Anualmente, alrededor de 100.000 invitados del extranjero realizan turismo en las áreas rurales gestionadas por estructuras comunitarias fuertes. Tanto el flujo de visitantes como la economía local han crecido anualmente un 13% y un 18%, respectivamente” (mincetur, n.d.).

“Como resultado del aporte turístico de a la zona rural económica, en el Perú se generaron 7 millones de nuevos soles” (mincetur, n.d.).

“Mediante el uso de recursos de personalización y basados en perfiles, que permitirán a los turistas especificar sus necesidades, las (TICS) pueden ayudar a los destinos a abordar con eficacia las necesidades de segmentos de mercado particulares” (Pereira et al., 2017).

“La sociedad moderna ha transformado fuertemente actualmente debido a los avances y desarrollos tecnológicos, haciéndola más compleja que nunca. Como resultado, se requieren

instrumentales y algoritmo de artificial inteligencia para equilibrar las preferencias del viajero. Por otro lado, la gran cantidad de información utilizada para administrar los negocios crece día a día, lo que requiere el uso de herramientas poderosas para extraer la información necesaria para realizar el análisis” (Rivasplata & Sahuanga, 2012). Uno de los instrumentos más populares para identificar paradigmas de conducta es la minería de información, y la instrucción automática tiene paradigmas de agrupación que pueden encontrar similitudes entre atributos y grupos de atributos que son diferentes entre sí.

“La característica principal de esta técnica es el uso de una métrica de similitud, que generalmente se basa en los caracteres que representan los objetos y normalmente se conceptualiza por la cercanía en un espacio en varias dimensiones.” (Rivasplata & Sahuanga, 2012).

El algoritmo de agrupamiento más utilizados es K-Means, que se determina por su simplicidad y tiene como propósito comprimir la cuantía de datos caracterizando o agrupando características similares. En este sentido, se debe utilizar un método de análisis de datos. Por ejemplo, el aprendizaje automático utilizado en turismo permitirá descubrir patrones de comportamiento turístico que el Ministerio de Turismo peruano, las agencias de viajes y los empresarios pueden utilizar para desarrollar estrategias para diferentes tipos de turistas. estrategia de mercado.

1.2. Objetivos de la investigación

1.2.1. Objetivo general.

Aplicar el algoritmo de k-means que permite clasificar los perfiles de turistas según la motivación de la TRC realizada en la región Puno, prediciendo su comportamiento futuro, busca mejorar y personalizar la atención.

1.2.2. Objetivos específicos.

- Adquisición, recopilación, integración, selección y depuración de información relacionada con los turistas que realizan CRT en la región Puno a través de encuestas.
- Implementar el algoritmo k-means en los datos recopilados de los turistas utilizando herramientas de aprendizaje automático.
- Valida el rendimiento del algoritmo de agrupamiento k-means para verificar que es una forma eficaz y útil de buscar fragmentos categóricos.

1.3. Justificación

Para identificar las características y similitudes entre la información que las agencias de viajes, corredores y empresarios recopilan sobre los turistas a través de encuestas, es necesario explorar conceptos relacionados. Por tanto, la existencia de una herramienta que nos permita entender y utilizar esta información nos ayudará a utilizar el aprendizaje automático (clustering) para conseguir nuestros objetivos e identificar segmentos del perfil turístico.

1.3.1. Social.

La investigación busca proporcionar información del turista que será útil a los Emprendedores, Municipios, Gobierno Regional y Mincetur para mejorar la atención al turista y dar una atención personalizada.

Identificar el perfil del turista y comprender el comportamiento, para brindarle una mejor atención, el cual permite al turista llevarse una buena imagen y una mejor experiencia del lugar que visito, permitirá la recomendación a otros turistas mediante buenos comentarios.

Generar empleo, optimar la existencia diaria de los individuos en zonas rurales que tienen buena cantidad de personas, generando oportunidades para que puedan iniciarse como

emprendedores de turismo, el cual la población no emigrara a las ciudades en busca de mejores trabajos.

Las comunidades se beneficien del desarrollo y que el entorno local se mantenga saludable.

1.3.2. Económico.

“El Empleo Turístico Directo e Indirecto en 2015 alcanzó los 1,3 millones. Según el Ministerio de Comercio Exterior y Turismo, entre 2011 y 2015 el PBI turístico aumentó en un monto acumulado mayor que el PBI nacional (Perú).” (Espinola, 2017).

Según Mincetur: “Alrededor de 100 millones de turistas extranjeros viajan anualmente a áreas rurales supervisadas por estructuras comunitarias de base. Hubo un aumento anual en el flujo de visitas del 18% y un aumento anual en la economía local del 13%.” (mincetur, n.d.).

Lograr una mayor afluencia de turistas no siendo un numero enorme, pero si con un perfil identificado, generará más trabajo y lograremos un beneficio directo para los hoteles, alojamientos, restaurantes y beneficios en la zona urbana y rural.

Entender las tipologías de los viajeros que ejecutan turismo rural en la región Puno, brindándoles una atención personalizada, aumentará la afluencia de turistas, generará aumento trabajos y una superior aptitud de vida para los empresarios, instituciones, municipios, traerá mejores ingresos al gobierno. región Perú.

1.3.3. Ambiental.

Recolección de datos en una aplicación móvil, porque no contamina el medio ambiente y no genera más basura, mediante la no utilización de hojas de papel para la recolección de datos, ayudando al desarrollo sostenible del turismo.

Al identificar el perfil del turista se contribuye con la sostenibilidad de recursos como el agua, el medio ambiente y otros recursos de las zonas que ofrecen turismo.

Generar políticas para el desarrollo del turismo sostenible (medio ambiente, empleo, económico).

1.3.4. Legal.

La ley número 30309 fomenta el avance tecnológico y la innovación tecnológica.

“Ley de Protección de Datos, LEY N° 29733 tiene por objeto hacer efectivo el derecho fundamental a la protección de datos personales consagrado en el artículo 2 Numeral 6 de la Constitución Política del Perú mediante el adecuado manejo de dicho derecho respetando los otros principales derechos reconocidos en dicha ley.” (Perú, n.d.)

1.3.5. Tecnológico.

Los turistas acceden y consumen productos turísticos de distintas formas, por lo cual ha generado nuevas necesidades y oportunidades para que podamos acceder a los datos, para una mejor comprensión del comportamiento.

Desde esta perspectiva, el turismo es ideal para desarrollar aplicaciones de recolección de datos (datos estructurados) y minería de datos, debido a que cualquier empresa o destino pueda capturar, monitorear, analizar e interpretar comportamientos de los turistas.

Para demostrar el potencial de la técnica de agrupación de k-means, en este caso, no hay grupos naturales en un grupo de turistas que viajan a través de comunidades rurales. Propiedades como el alto potencial de uso en datos de alta dimensión y la facilidad de lectura de los resultados justifican su elección entre otros métodos de agrupamiento.

1.4. Presunción filosófica

Un corazón sabio adquiere conocimiento, y un oído perspicaz lo busca. Proverbios 18:15.

La Educación en la Universidad Peruana Unión es un desarrollo armónico de las facultades

Preparar a los estudiantes física, mental y espiritualmente para el servicio en el mundo.

El verdadero conocimiento y su desarrollo provienen del reconocimiento de Dios.

“Un maestro genuino no estará satisfecho con un trabajo mediocre. No se siente cómoda guiando a sus alumnos hacia una meta inferior a la más alta que son capaces de alcanzar. No es suficiente simplemente impartirles conocimientos técnicos o convertirlos en comerciantes exitosos, contadores experimentados o artistas veteranos. Su objetivo es inculcarles los valores de la verdad, la obediencia, el honor, la integridad y la pureza, que los convertirán en fuertes fuerzas para la elevación y estabilidad de la sociedad. Espero especialmente que aprendan la lección más importante de la vida: la lección del servicio traicionado” (White, 2009)

Capítulo II. Marco Teórico

2.1. Revisión de la literatura

Tabla1.
Revisión de la literatura

Autor	Año	Título	Objetivos	Método	Resultados
Manish Verma			Se compara un conjunto de datos bancarios relacionados con la información bancaria utilizando diferentes métodos de segmentación para determinar las fortalezas y debilidades de cada método y qué método es más efectivo.	La herramienta donde se procesó la información es el software Libre Weka.	(1) "Cuando el error cuadrático medio disminuye, el rendimiento del algoritmo k-means aumenta y, a medida que aumenta el número de conglomerados, el error cuadrático medio disminuye". (2) "El rendimiento del Algoritmo k-
Maully Srivastava		"A Comparative Study of Various Clustering Algorithms in Data Mining"			
Neha Chack	2012				
Atul Kumar Diswar					
Nidhi Gupta					

				medias es mejor que el Algoritmo de Segmentación Jerárquico”. (3) “El Algoritmo K-medias es más eficaz que los demás algoritmos y a su vez efectúa conclusiones de alto impacto con datos extensos”.	
Franco Rafael Bazán Pérez	2015	“Puesta en práctica de la técnica de segmentación estadística Clúster K-medias para comprender las necesidades financieras de los clientes de las financieras”.	Mostró el gran potencial de la técnica de clustering k-means, que suple la falta de poblaciones naturales en la clientela de las entidades financieras.	Utilice la técnica de agrupamiento de k-medias, es decir, la metodología de la técnica, si el consenso se ha mostrado favorable al tipo de	Los resultados de aplicar la técnica estadística de clustering de K-means basada en

información que posee.	información histórica de los clientes durante al menos 1 año pueden confirmar el alto potencial de la técnica, por lo que se ofrecen 7 grupos de clientes diferentes, que son: Anclada, Consolidada, Prosperidad, Consolidada, Agresivo, Consolidado, Madurez, Consolidada, Ambición, Indifere
------------------------	--

					nte, Ambición conservadora y Ambición madura. Deseos apáticos , deseos conservadores y deseos maduros. Dentro de cada grupo, agrupamos a los clientes en función de la demografía (estudiantes, trabajadores, jubilados, agricultores, ciudadanos acaudalados) y utilizamos variables RFM
Varun Kumar. M				Identificar grupos de clientes con necesidades comunes para mejorar las estrategias de marketing.	La metodología de valoración de clientes RFM.
Vishnu Chaitanya. M	2012	Segmenting the Banking Market Strategy by Clustering.			El enfoque LTV (valor de vida del cliente) le permite generar perfiles de clientes.
Madhavan. M					

				(depósitos, inversiones, transacciones) como promedios centrales.
				Los resultados por medio de los análisis permitieron identificar dos conglomerados claramente definidos: El clúster 1 agrupa a estudiantes con alto rendimiento académico que demuestran un mayor nivel socioeconómico,
Alicia Alelí Chaparro Caso López				
Coral González Barbera	2016	“configuración de perfiles de la Familia y rendimiento académico de estudiante de secundaria”	Creación de perfiles de estudiantes de secundaria en función del rendimiento académico, el nivel socioeconómico, el capital cultural y la organización familiar.	
Joaquín Caso Niebla				

mayores
indicad
ores de
capital
cultural
y mayor
particip
ación
familiar
;
mientra
s que
los
grupos
del
conglo
merado
2 tienen
menor
rendimi
ento
académi
co,
nivel
socioec
onómic
o y
capital
cultural
Los
estudian
tes con
puntajes
más
bajos y
menor
particip
ación en
las
organiz
aciones
familiar
es se

					agrupan .
					Confor ma la data de los encuest ados utilizad os por los investig adores, la capital percibe viajeros de todas las clases socioec onómic as, pero el Nivel 3 y el Nivel 4 domina n con un 63 %, y la mayoría son emplead os o autóno mos con un 67 %. El 65% de los usuarios no utiliza
Duván Emilio Ramírez Opina.				Identificar los aspectos más importantes que influyen en el comportamient o de los consumidores de viajes y su influencia en la decisión de compra.	
María Cristina Otero Gómez.	2014	“Comportamie nto del consumidor Turístico”.		Metodologí as cualitativa y cuantitativa	
Wilson Giraldo Pérez					

Kenny Alejandr o	2014	Descripción de perfiles de clientes en el área de	Diseñar y fomentar un instrumento basada en la	Proceso estándar interdiscipli nario para	Internet para elegir destino, el 47% de sus fuentes de informa ción les son conocid as y el 43% de los consumi dores llegan a un conjunt o de amigos; Un dato destaca ble es que el 37% de los consumi dores encuest ados vivía con familiar es y el 18% vivía con amigos. Los concluy entes derivad
------------------------	------	--	---	--	---

Morelo Tapias.	negocio de la mype franja t.	web que permita la extracción de datos para perfilar a los consumidores comerciales de la zona T y pronosticar su comportamiento futuro; tratar de mejorar y ajustar su cuidado.	la extracción de datos (CRISP-DM).	os admitieron asemejar tres (3) grupos distintos de clientelas, enfocándose en las diligencias de marketing y servicios al usuario, que fueron los principales objetivos del proyecto.
----------------	------------------------------	--	------------------------------------	--

Fuente: Elaboración propia.

2.2. Marco teórico

La base teórica de este estudio incluye: turismo, segmentación de mercado, desarrollo de aplicaciones y minería de datos.

2.2.1. Turismo.

“Debido a las numerosas interpretaciones de este fenómeno social, la conceptualización del turismo ha dado lugar a importantes controversias. Estas interpretaciones son dadas por diversas disciplinas que, de alguna manera, incluyen el turismo en sus investigaciones, dando como resultado definiciones parciales del turismo. Un ejemplo es el debate sobre si el turismo es un servicio o una industria. Se brindan interpretaciones desde varias perspectivas, pero dan una idea de lo que representa el turismo para varias disciplinas, por ejemplo” (Farro, 2010).

- Por economía: el turismo es un servicio.
- Para las ciencias geográficas: el turismo es el movimiento de personas de un punto a otro sobre la superficie de la tierra.
- Para la ciencia política: el turismo es el ejercicio de los derechos propios del individuo a la libertad de tránsito.
- Para la sociología: viajar es una oportunidad para satisfacer necesidades que suelen dejarse de lado, cuando un individuo se dedica a su vida laboral.

Como se puede ver, definir el turismo es una tarea difícil. Según Acerenza, el turismo tiene un significado propio desde una perspectiva social: "La esencia del turismo para la sociedad es una actividad relacionada con el placer, la relajación, la promoción cultural y la recreación" (Acerenza, 2006).

“Las motivaciones del turismo como actividad humana varían en función de los deseos de cada individuo como resultado. En este sentido, la motivación es un proceso dinámico que cambia según la experiencia, junto con otros factores como la edad y el estatus” (Ramírez, Otero, & Giraldo, 2014).

2.2.1.1. Turismo Rural Comunitario.

“Todas las actividades relacionadas con el turismo que se desarrollan en las zonas rurales del Perú deben planificarse con anticipación y deben ser sostenibles. También deben basarse en la colaboración de los ciudadanos del local organizado en merced de los pobladores, con la cultura rural como dispositivo fundamental.” (Mincetur, s.f.).

2.2.1.2. Segmentación de mercado turístico.

Cada destino turístico solo puede atender adecuadamente un tipo específico de necesidad porque ofrece una gama de recursos que atraen a segmentos de mercado específicos porque los mercados no son uniformes. Por lo tanto, las tácticas de marketing deben tener en cuenta de motivación de viaje de los turistas de hoy para desarrollar ofertas que se adapten a las miserias únicas del propósito del mercado. Por lo tanto, la segmentación es fundamental para brindar productos y servicios que se adapten a cada segmento como ya se mencionó, Wendell R. Smith lo definió como “el proceso de dividir el mercado en subgrupos homogéneos con el fin de llevar a cabo una estrategia comercial diferenciada para cada uno de ellos, que les permita satisfacer más eficazmente sus necesidades y lograr los objetivos comerciales de la empresa” en 1956” (Ferreira et al., 2009).

“Desde la concepción y dada su trascendencia en el campo del marketing, se han realizado numerosos estudios en este ámbito. Como resultado, lo que comenzó como una definición teórica se puso en marcha muchas industrias, incluida la turística, al buscar y establecer características que se aplicaran a grupos o segmentos homogéneos.” (Meiriño, Fraiz Brea, Araújo Vila, & Rivo López, 2016).

2.2.2. Inteligencia Artificial.

Imitar una parte de la inteligencia humana en sistemas artificiales es el objetivo del campo de la inteligencia artificial (IA), que está conectado con la teoría computacional. Los usos más comunes de IA son el procesamiento de datos y el reconocimiento del sistema porque la inteligencia humana se define únicamente como procesos de percepción sensorial (visión, audición, etc.) y procesos de reconocimiento de patrones relacionados.

“La creación de un sistema de inteligencia artificial suele requerir el uso de herramientas de disciplinas muy diferentes, como la computación matemática, la estadística, las tecnologías de la información, el procesamiento de señales, el control automático, la robótica o la neurociencia. Por ello, a pesar de que la artificial inteligencia se considera una sub categoría de la automatización teórica, es un campo al que contribuyen activamente muchos científicos, técnicos y matemáticos. En algunos casos, también es beneficiosa la indagación en otras áreas como la psicología, la sociología o la filosofía.” (Benítez, Escudero, & Kanaan, 2014)

2.2.2.1. *Machine Learning.*

“Es una técnica que ayuda a que el software o las computadoras aprendan de los datos sin programación explícita y se enfoca en hacer predicciones usando colecciones de "Datos de Entrenamiento". Con esta técnica, es posible distinguir entre el correo electrónico no deseado y no deseado, conocer las preferencias de los usuarios y hacer recomendaciones basadas en ellas, desarrollar estrategias para atraer clientes potenciales y más.” (Mérida & Ríos, 2014)

2.2.2.2. *Deep Learning.*

“El aprendizaje profundo, también conocido como aprendizaje profundo, es una técnica de codificación relativamente nuevo que se cimienta en redes de neuronas corticales.” (Arista, Calderón, Fierro, & Nakano, 2017)

2.2.3. Técnicas Descriptivas o No supervisadas.

En los métodos de descripción, a las variables no se les asignan roles predefinidos. No se asume ni la variable dependiente ni las variables independientes ni la coexistencia de un prototipo a priori de la data. Los patrones se ejecutan en automático en función del reconocer sus patrones

2.2.3.1. Clustering o Agrupamiento.

Divide la información en diferentes grupos. El objetivo de la agrupación es encontrar grupos (clúster) que sean muy diferentes entre sí y cuyos miembros sean muy similares entre sí. A diferencia de la clasificación, cuando comienza, no sabe qué grupos aparecerán o qué tipo de atributos se agruparán los datos. Por tanto, estos grupos deben ser interpretados por una persona con conocimientos empresariales. Los grupos o conglomerados identificados se pueden utilizar para clasificar nuevos datos. Algoritmo de K-Means:

“Creado en 1967 por Mac-Queen. Este algoritmo es muy apreciado porque es simple de implementar y porque la complejidad del algoritmo depende de la cantidad de objetos. El inconveniente principal con este algoritmo es que es muy sensitivo a la parte elegida inicialmente y puede convergir corridamente. Sin embargo, es el más acreditado y ampliamente manejado porque es muy fácil de aplicar y efectivo. Siga un proceso sencillo para categorizar un grupo de elementos en un número predeterminado de clúster basados en prioridades. K-Means representa cada conglomerado por la mediana de sus puntos, o su centroide, y cada conglomerado se caracteriza por esto, que se encuentra en el centro de las partes constituyentes del conglomerado.” (Rivasplata & Sahuanga, 2012).

2.2.3.2. Funcionamiento del Algoritmo k-means.

K-means es transcrito como K-medias y se realiza en las siguientes fases esenciales:

- Etapa 1: elegir al azar K objetos que constituyen los K clúster iniciales. El valor central inicial para cada conglomerado k es x_i , por lo que este es el único que pertenece a los conglomerados.
- Etapa 2: Se requiere reasignar los elementos del roster. Según una medida de distancia, el prototipo asignado a cada objeto x es el que está más cerca del objeto (habitualmente la medida euclidiana).
- Etapa 3: Después de que todos los elementos hayan sido ordenados, calcule los puntos centrales del K -Clúster. Estos nuevos clústeres son el promedio de todos los objetos asignados al clúster.
- Etapa 4: Repetir las etapas 2 y 3 hasta que no ocurran más reasignaciones. Aunque el algoritmo siempre tiene éxito, no se garantiza obtener la solución ideal. “Al final, el algoritmo es muy sensible a la selección aleatoria de los K centros iniciales. Por esta razón, el algoritmo K -Means se aplica varias veces al mismo conjunto de datos en un esfuerzo por minimizar este efecto. Esto se debe a que se sabe que los centros que inicialmente están lo más abiertos posible producen los mejores resultados.” (Alberto & Ramos, 2016).

La determinación de K es la siguiente:

- Si K es muy pequeño, se agruparán grupos “distintos”.
- Si se elige un K muy grande, hay centros que pueden quedar huérfanos, o sin agrupación.
- El valor de K puede determinarse según alguna heurística.

Por tanto, para conseguir un valor de K óptimo o una aproximación concluyente, se decidirá realizar numerosos experimentos con los datos para, tras analizar los resultados, conseguir una mejor estimación de la variable K.

2.2.3.3. *Distancia Euclidiana.*

El método más popular y fácil de entender es euclidiano. Las n-uplas ordenadas por n números reales son las componentes de Sea R_n , un espacio euclidiano de n dimensiones. Datos de objeto I1 e I2 medidos según n variables. (X_1, \dots, X_n) :

$$d_{I_1 I_2} = \sqrt{\sum_{k=1}^n (X_{1k} - X_{2k})^2}$$

2.2.3.4. *Distancia Minkowski*

Generalización de la distancia Euclidiana mediante el parámetro r.

$$dist = \left(\sum_{k=1}^n |p_k - q_k|^r \right)^{\frac{1}{r}}$$

r = 1. Distancia Manhattan

Ejemplo típico: Distancia de Hamming: Numero de bits diferentes entre dos arreglos de bits λ

r = 2. Distancia Euclidiana λ r $\rightarrow \infty$. Distancia “supremo” (norma L_{\max} o L_{∞}).

2.2.3.5. *Distancia Mahalanobis.*

$$d_{ABS}(x, y) = \sqrt{\sum_{i=1}^n |x_i - y_i|}$$

Esta distancia se usa comúnmente cuando los datos provienen de una o varias ciudades, pero con una matriz de varianza-covarianza alta. Σ , común.

2.2.4. Variantes del Algoritmo K-means.

2.2.4.1. K-medoides.

“Es un algoritmo de agrupación único que difiere ligeramente del algoritmo k-means. El algoritmo k-means selecciona los medios como los centroides, pero el algoritmo K-medoides selecciona los puntos de datos originales como los medoides. Un medoide se define como un objeto en un grupo cuya disimilitud promedio a todos los demás objetos en el grupo es la más pequeña. Cada objeto restante se agrupa con el medoide más cercano, y estos algoritmos realizan iterativamente todos los intercambios posibles entre los objetos representativos y los que no lo son, hasta que se minimiza una medida de disimilitud entre los k-medoides y los vectores de observación que forman los conglomerados.” (Chamba, 2015).

2.2.5. CRISP-DM.

Es un estándar para la ejecución de proyectos de minería de datos, con el ciclo de un proceso de minería de datos dividido en seis etapas. Algunos de estos pasos están bifurcados, lo que significa que algunos de ellos le permitirán volver atrás y revisar parcial o completamente los pasos anteriores:

Entendimiento del negocio, Comprensión de datos, Recopilación y preparación de datos Modelado, Evaluación e Implementación.

Su objetivo es completar proyectos de minería de datos que sean menos costosos, más confiables, repetibles, manejables y ágiles.

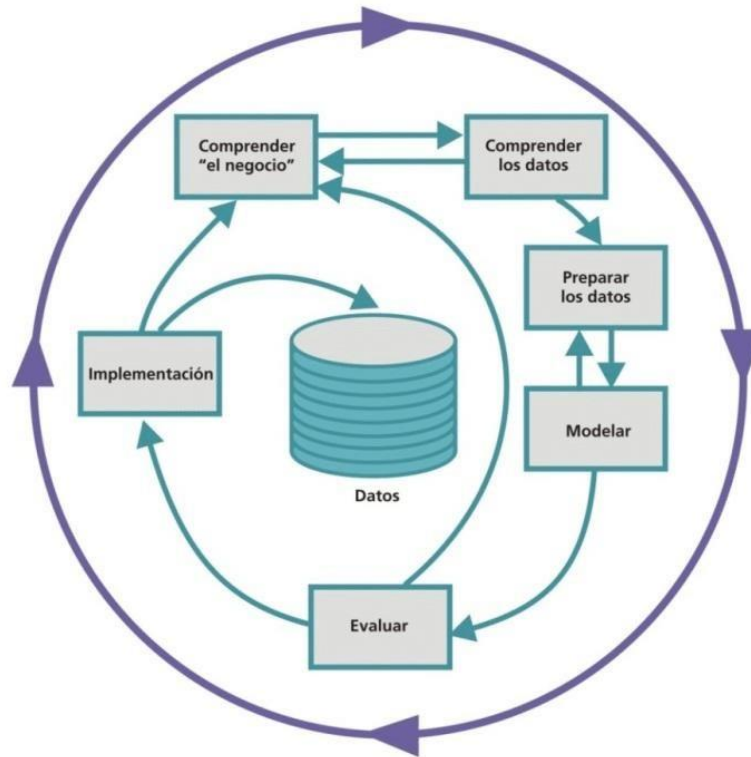


Figura 1. Proceso de Metodología ágil CRISP-DM

2.2.5.1. Fase de comprender el negocio.

La tarea de comprender los propósitos y exigencias del proyecto desde una representación empresarialmente o gubernamental y traducirlos en propósitos tecnificados y planes proyectados es probablemente la más crucial y la que requiere más tiempo.

2.2.5.2. Fase de comprensión de los datos.

La recogida inicial de datos forma parte de la segunda fase, que tiene como objetivo establecer un primer punto de contacto con el tema, familiarizarse con él, evaluar su calidad y establecer las conexiones más claras que permitan definir las hipótesis iniciales. En un proyecto para una instalación de minería de datos, esta fase y las dos siguientes demandan la mayor cantidad de esfuerzo y tiempo (MD). En términos generales, si una organización ya cuenta con una base de datos corporativa, es preferible desarrollar una nueva base de datos ad-hoc para el proyecto MD, ya que durante el desarrollo del proyecto puede surgir un acceso frecuente y abundante a la base

de datos corporativa, lo que lleva a consultas. y posiblemente modificación, lo que podría generar muchos problemas.

2.2.5.3. Fase de preparación de los datos.

Para familiarizarse con los datos teniendo en cuenta los objetivos del negocio, se deben completar las siguientes tareas. recopilación inicial de datos, descripción de los datos, exploración de datos y aseguramiento de la calidad de los datos una etapa en el proceso de preparación de datos. Esta etapa vio el desarrollo de la selección, limpieza e integración de datos.

2.2.5.4. Fase de modelado.

Durante esta fase, se seleccionará la técnica de modelado, seguida del diseño de evaluación, construcción del modelo y evaluación del modelo.

2.2.5.5. Fase de evaluación.

Aquí se realizó la evaluación de los modelos de las fases anteriores para determinar si son útiles a las necesidades del negocio y debe considerarse lo siguiente: Evaluación de resultados y revisar el proceso.

2.2.5.6. Fase de implementación.

Una vez creado y validado el modelo, el conocimiento adquirido se traducirá en acciones dentro del proceso de negocio, ya sea que el analista sugiera acciones basadas en la observación del modelo y sus resultados, ya sea aplicando el modelo a varios conjuntos de datos o como parte del proceso.

2.2.6. Herramientas de Machine Learning.

Tabla 2.
Herramientas para Machine Learning

Herramienta	Descripción	Arquitectura
--------------------	--------------------	---------------------

Weka	Es una colección de algoritmos de aprendizaje automático para tareas de minería de datos. Contiene herramientas para la preparación de datos, clasificación, regresión, clustering, minería de reglas de asociación y visualización. Weka es un software de código abierto emitido bajo la Licencia Pública General de GNU.	Java
Graphlab	Turi es un framework de cálculo distribuido basado en gráficos, de alto rendimiento, escrito en C ++. Es un proyecto de código abierto que utiliza una licencia Apache. Permite desarrollo simple de modelos personalizados de Machine Learning.	Python Sframe SGrapph Machine Learning Python Numpy y Scipy Matplotlib
Scikit-learn	Es una biblioteca de Machine Learning, software libre para el lenguaje de programación.	R
R	Es un entorno de software libre para computación estadística y gráficos.	R

Fuente: Elaboración propia

2.2.7. Limpieza de datos.

2.2.7.1. Python.

Python es fácil de usar, pero es un verdadero lenguaje de programación que brinda mucha más estructura y soporte para programas grandes que los scripts de Unix o los archivos por lotes. Por otro lado, Python proporciona mucha más verificación de errores que C y, como lenguaje de alto nivel, incluye estructuras de datos de alto nivel, como matrices y diccionarios de tamaño flexible. Debido a sus tipos de datos más generales, Python se puede aplicar a una gama más amplia de problemas que Awk o incluso Perl, y muchas cosas siguen siendo al menos tan simples en Python como lo son en esos lenguajes.

Python permite la creación de programas compactos y legibles. Por diversas razones, los programas de Python suelen ser más cortos que sus equivalentes en C, C++ o Java.

- No es necesario declarar variables o argumentos porque las operaciones complejas se pueden expresar en una sola instrucción gracias a los tipos de datos de alto nivel.
- La agrupación de instrucciones también se realiza a través de la sangría en lugar de abrir y cerrar cerraduras.

“Python es extensible: si está familiarizado con la programación en C, es simple agregar una nueva función o módulo al núcleo, ya sea para realizar operaciones críticas a la velocidad más rápida posible o para integrar programas de Python con bibliotecas que solo pueden ser disponibles en forma binaria (por ejemplo, bibliotecas de gráficos específicas del fabricante). Una vez que esté realmente entusiasmado, puede incrustar Python en una aplicación C y usarlo como lenguaje de extensión o lenguaje de comando de esa aplicación.” (Van, 2013).

Capítulo III. Materiales y Métodos

3.1. Descripción del lugar de ejecución.

El proyecto se desarrolla en el distrito de Juliaca ubicado a 3822 m.s.n.m. Provincia de San Román, Departamento de Puno en las instalaciones de la oficina de CONCYTEC que tiene la Universidad Peruana Unión Filial – Juliaca. El centro donde se pretende obtener los datos del turista será del distrito de Amantani, mediante los emprendedores que brindan el servicio del TRC.

3.2. Materiales e insumos

Para levantar información de los turistas se realizó durante periodos asignados de PromPerú que fueron necesario elaborar 1 formulario de tipo encuesta, referencia de PromPerú. Esta encuesta se desarrolló de acuerdo a las variables necesarias para la recolección de datos del turista que realiza TRC. Los datos fueron recopilados mediante las encuestas físicas.

Las herramientas para el desarrollo de la investigación:

- Laptop TOSHIBA Disco duro de 1 TB, memoria RAM 16 Gb, procesador i7 2.60GHz, pantalla 15". Computador donde se desarrolló la app móvil, web y se estableció el algoritmo.
- Python 3.5. Lenguaje de programación ágil.
- PyCharm 2017.2.3. Contexto de progreso compuesto para Python.
- Anaconda. Entorno de proceso para machine learning en Python.

3.3. Metodología

3.3.1. Tipo de la Investigación.

3.3.2. Arquitectura de solución.

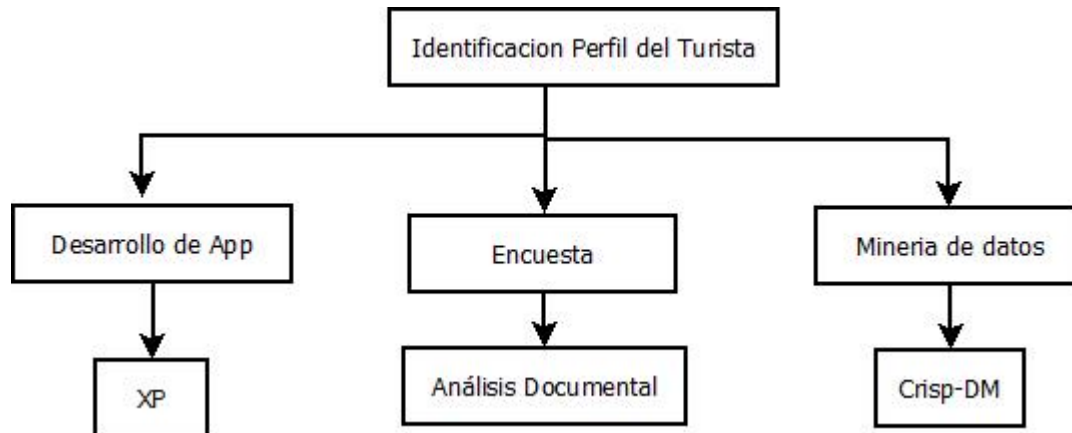


Figura 2. Metodología para la identificación de perfil del turista

3.3.3. Encuesta.

Análisis documental: Consistió en la revisión de documentos, artículos, tesis, libros, internet, que contenían información acerca del tema de la investigación Cuestionario de PromPeru.

- Tema de interés: Turismo Rural Comunitario
- Objetivo: Conocer las motivaciones del turista que optó para realizar TRC.
- Preguntas: anexo Encuesta PromPeru
- Población: Turistas en la región Puno.
- Periodo: 1 – 3 meses.
- Validación de encuesta:

La aplicación del instrumento estará a cargo de los encuestadores, mediante la encuesta, lo cual ayudo a recoger la data conforme los periodos realizados por PromPeru).

3.3.4. Programación Extrema(XP).

El método XP, también conocido como Programación Extrema, es un método de gestión de proyectos rápido y flexible.

Extreme Programming se enfoca en mejorar las relaciones interpersonales del equipo de desarrollo como clave para el éxito a través del trabajo en equipo, el aprendizaje continuo y un ambiente de trabajo positivo.

Esta metodología se enfoca en la retroalimentación continua entre el cliente y el equipo de desarrollo y es ideal para proyectos con requisitos imprecisos y que cambian rápidamente.

3.3.5. CRISP-DM.

En el caso de los proyectos de implementación de minería de datos, hay uno en particular; CRISP-DM, que recibió apoyo de empresas privadas y organismos públicos en sus primeros años de publicación.

La norma incluye un modelo y una guía que se dividen en seis etapas. Algunas de las etapas son bidireccionales, lo que significa que algunas de las etapas le permitirán repasar parcial o totalmente las etapas anteriores.

Este método para proyectos de minería de datos no es el "más actual" ni el "mejor", pero es muy útil para entender la tecnología y generar ideas para diseñar o revisar métodos de trabajo para proyectos de características similares.

3.4. Cronograma del proyecto

Código	Nombre	Días	Fecha Inicio	Fecha Fin
1.1	Elaboración de encuesta (Instrumento) y recolección de datos	90d.		
1.2	Proceso de minería de datos con k-means y herramientas	60d		
1.3	Redacción de informe de tesis	60d		

3.5. Presupuesto

ITEM	DESCRIPCIÓN	UNIDA	CANTIDAD	PRECIO S/.	PARCIAL S/.
01.02	BIENES				0.00
01.02.01	Material impreso de libros electrónicos	Unid	4.00	10.00	40.00
01.02.02	Adquisición de libros de consulta	Unid	10.00	12.00	120.00
01.03	RR.HH				0.00
01.03.02	Analista	Mes	1.00	1100.00	1100.00
01.03.03	Programador	Mes	2.00	1000.00	2000.00
01.04	SERVICIOS				0.00
01.04.01	Servicio de internet	Mes	4.00	120.00	480.00
01.04.02	Viáticos y Movilidad	Mes	1.00	500.00	500.00
01.05	MATERIALES				0.00
01.05.01	Tablet	Unid	1.00	1500.00	1500.00
01.05.03	Impresora Láser color	Unid	1.00	200.00	200.00
01.05.05	Celular	Unid	2.00	400.00	800.00
01.05.06	Papel bond A4	Mllr	1.00	12.50	12.00
01.05.07	Bolígrafo azul	Unid	5.00	2.00	10.00
01.05.08	Marcador color amarillo	Unid	3.00	3.00	9.00
01.05.09	Clips	Cja	1.00	4.00	4.00
01.05.10	Folders manila A4	Pqte	1.00	40.00	40.00
01.05.11	Engrapador alicate	Unid	1.00	35.00	35.00
01.05.12	Grapas para engrapador alicate	Cja	2.00	5.00	10.00
01.05.13	Perforador	Unid	1.00	10.00	10.00
01.05.14	Thonner(impresora)	Unid	1.00	250.00	250.00
01.06	Gastos imprevistos				210.00
PRES.TOTAL					7330.00

EL proyecto es financiado por el Proyecto, UPEU -CONCYTEC.

Capítulo IV. Resultados y discusión

Tras 9 iteraciones se alcanzó la solución final con 1122 datos, En el clúster 1 y 2 podemos apreciar que los datos nos dan la referencia, muestra 2 segmentos de perfil de turista.

El clúster 1 se puede identificar como el perfil para el Turismo Rural Comunitario con la siguiente interpretación, Lugar de salida Lima, Hacer Camino Inka, Conocer/visitar comunidades nativas/ andinas / campesinas, ingresaron Puno (Desaguadero), ingreso por Primera Vez al Perú, Contraté todos los servicios y actividades en un paquete, Grupo de viaje Con mi pareja, Compran su paquete en el Mes de Junio, Deportes / Deportes de aventura

En la Tabla 3 se muestra que variables contribuyeron a la solución de los conglomerados.

Tabla 3.
Variables

LUGAR	Lugar de Salida:
P02_1	¿Qué aspectos fueron los que más lo motivaron a visitar el Perú en este viaje?
P03	De todos los aspectos mencionados ¿cuál diría usted que fue el motivo más importante?
P06	¿Por dónde ingresó al Perú?
P07	¿Es ésta la primera vez que usted visita el Perú, o lo ha visitado más de una vez?
P09Pais	En qué país reside usted?
P17A	¿Cómo organizó su viaje?
P21_GrupoDeViaje	¿Cómo organizó su viaje?
P35A	¿En qué mes compró su pasaje y /o paquete para realizar este viaje?
P36A_1	Usualmente. ¿Cuáles aspectos que más influyen en Ud. para elegir un destino
P37A_1	¿Qué medios son los que más han influido en Ud. para decidir un destino?
P44_20	En total, ¿Cuántas noches pasó en el departamento de..., visitando...? – Puno
P45_20_1	¿En qué tipo de alojamiento pernoctó en el Departamento de...? – Puno
P54	¿Cuál de las frases que le voy a mostrar se identificaría más con usted?
P62	¿Qué edad tiene usted?
P63	Género
P64A	Según la siguiente tarjeta, ¿Cuál es su condición actual?
P66	¿En cuál de las siguientes categorías se encuentra actualmente?
P68	Grado de instrucción más alto concluido

```

Final cluster centroids:

```

Attribute	Full Data (1122.0)	Cluster#			
		0 (237.0)	1 (433.0)	2 (284.0)	3 (168.0)
LUGAR	1.2629	2	1.0162	1.1761	1.006
P02_1	12.5089	6.2236	14.6674	14.4718	12.494
P03	4.0909	3.1181	4.6813	3.3908	5.125
P06	1.1818	1.3038	1.1039	1.088	1.369
P07	1.0383	1.0169	1.037	1.0528	1.0476
P09Pais	148.7264	100.8354	169.7483	83.5317	272.3155
P17A	1.6711	1.2194	1.7875	2.1761	1.1548
P21_GrupoDeViaje	2.1604	2.0169	2.3095	2.2676	1.7976
P35A	5.8939	5.6203	6.2656	5.2148	6.4702
P36A_1	11.6301	1.865	20.4573	7.6232	9.4286
P37A_1	9.0936	6.4008	10.8822	8.7007	8.9464
P44_20	2.0116	1.7679	2.1617	2.0528	1.8988
P45_20_1	3.9323	4.7257	3.3326	2.6725	6.4881
P54	1.1016	1.1435	1.1178	1.0423	1.1012
P62	40.2977	30.4473	43.7829	47.2711	33.4226
P63	1.4831	1.4599	2	1	1
P64A	2.1168	1.7468	2.2402	2.2289	2.131
P66	20.3271	6.3924	24.2448	25.9296	20.4167
P68	4.4144	4.4726	4.388	4.6655	3.9762
P69	79.5677	26.9662	93.5035	95.0704	91.6488

Figura 3. Variables Definidas para el Análisis

En la figura 4 Este método hace uso de los valores de inercia obtenidos tras aplicar las K-medias a varios números de Clúster (desde 1 a N Clúster), siendo la inercia la suma de las distancias al cuadrado de cada objeto del Clúster a su centroide, permitió poder elegir el número de clúster:

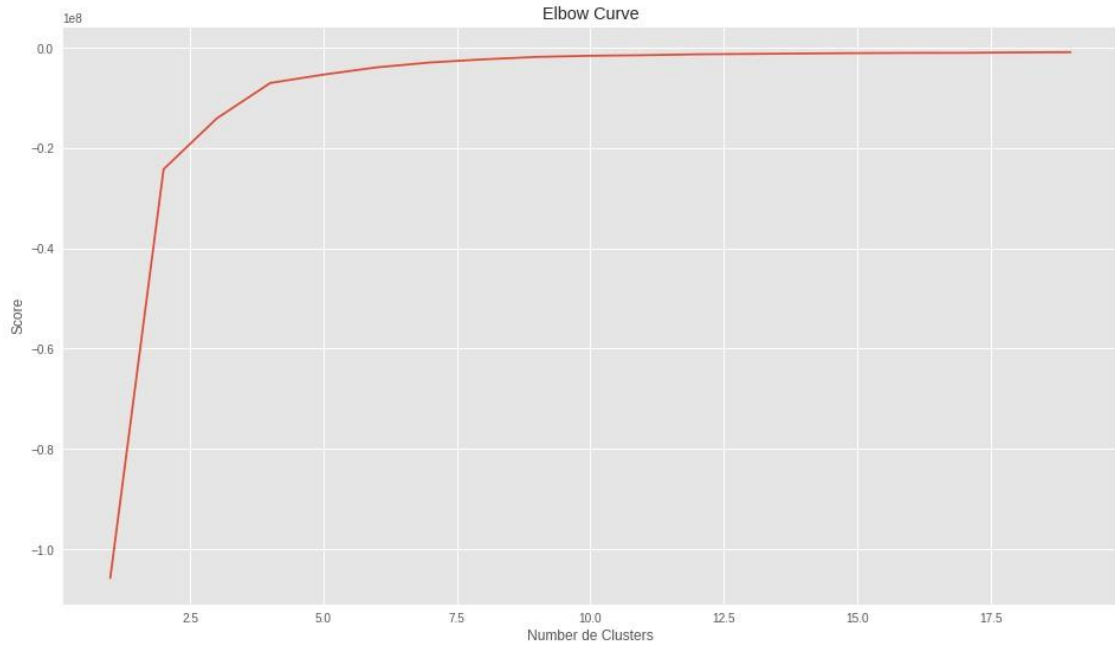


Figura 4. Grafica del punto de codo para el numero de clúster.

En la figura 5 nos muestra gráficamente la visualización de los 4 clúster en distintos colores debidamente identificados por medio de la herramienta weka.

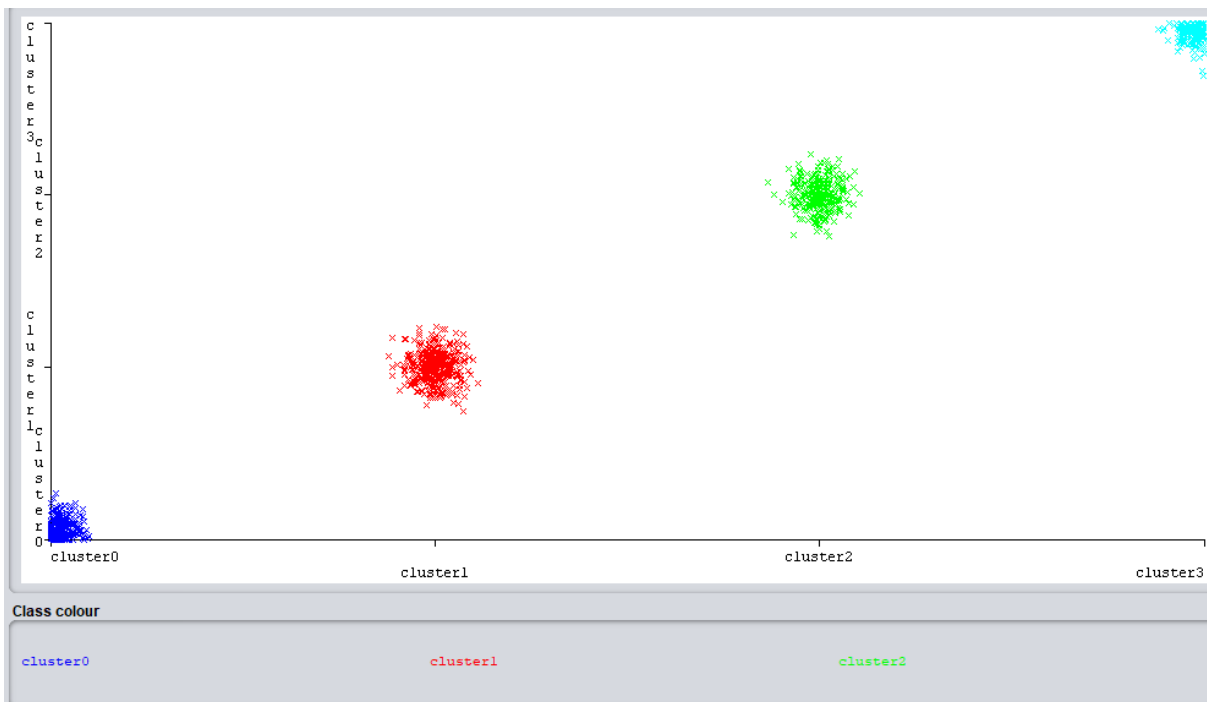


Figura 5. Clústeres definidos gráficamente.

Capítulo V. Conclusiones y Recomendaciones

Como metodología de segmentación, logró validar el potencial de la técnica de agrupamiento de k-means, brindando una solución óptima a la ausencia de perfiles naturales entre los turistas que realizan turismo al postular dos segmentos distintos.

Se confirma que la técnica de agrupamiento de k-medias tiene un alto potencial de segmentación como lo demuestra su desempeño con una muestra y la población general, así como la clasificación de los resultados de la observación. En consecuencia, también es acertado afirmar que esta técnica funciona de forma óptima a la hora de encontrar diferencias sustantivas entre los turistas del estudio.

Con base en los datos presentados en este estudio, se puede afirmar que las variables pertenecientes al clúster 1 y 2 para el segmento del perfil del turista, resulto apropiado para la configuración del perfil del turista.


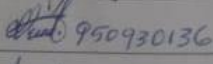
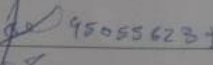
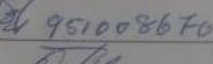
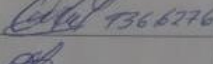
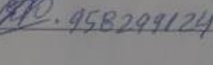
REFERENCIAS BIBLIOGRÁFICAS

- Acerenza, M. Á. (2006). *Conceptualización, Origen Y Evolución Del Turismo*.
<https://doi.org/380.1459104'A545c>
- Alberto, F., & Ramos, C. (2016). *Clasificación automática de Tweets utilizando K-NN y K-Means como algoritmos de clasificación automática, aplicando TF-IDF y TF-RFL para las ponderaciones Profesor Guía : Rodrigo Alfaro Arancibia*.
- Arista, A., Calderón, G., Fierro, A., & Nakano, M. (2017). Clasificación de imágenes urbanas aéreas: Comparación entre descriptores de bajo nivel y aprendizaje profundo. *Información Tecnológica*, 28(3), 209–224. <https://doi.org/10.4067/S0718-07642017000300021>
- Benítez, R., Escudero, G., & Kanaan, S. (2014). *Inteligencia artificial avanzada*. (E. de la U. O. de Catalunya, Ed.). Catalunya, España.
- Chamba, S. (2015). *Minería de Datos para segmentación de clientes en la empresa tecnológica Master PC*. (Tesis de pregrado) Universidad Nacional de Loja. Loja, Ecuador.
- Espinola, A. E. (2017). *Importancia de la cuenta satélite de turismo en el Perú*.
- Farro, V. D. (2010). *APLICACIÓN DE LAS CADENAS DE MARKOV EN LA DETERMINACIÓN DE CIRCUITOS TURÍSTICOS DEL PERÚ*. (Tesis de pregrado) Pontificia Universidad Católica del Perú. Lima, Perú.
- Ferreira, S. D., Rial, A., & Varela, J. (2009). Post Hoc Tourist Segmentation with Conjoint and Cluster Analysis. *Pasos*, 7, 491–501.
- Ferreira, S. D., Rial, A., & Varela, J. (2010). SEGMENTACIÓN POST HOC DEL MERCADO TURÍSTICO ESPAÑOL Aplicación del análisis clúster en dos etapas, *19*, 43–48.
- Guterres, A. (2015). *Organización Mundial del Turismo. UNWTO Annual Report 2015*. Retrieved from <http://www2.unwto.org/es/annual-reports>

- Meiriño, R., Fraiz Brea, J. A., Araújo Vila, N., & Rivo López, E. (2016). Segmentación del mercado de un destino turístico de interior. El caso de A Ribeira Sacra (Ourense). *PASOS. Revista de Turismo y Patrimonio Cultural*, 14(2), 369–383. Retrieved from <http://www.redalyc.org/articulo.oa?id=88145251005>
- Mérida, C. M., & Ríos, R. P. (2014). *Propuesta de la plataforma de big data orientado al sector turístico (Tesis de Pregrado)*. Universidad Peruana de Ciencias Aplicadas.
- mincetur. (n.d.). Turismo Rural Comunitario - Mincetur. Retrieved September 18, 2018, from <https://www.mincetur.gob.pe/producto-turistico/turismo-rural-comunitario/>
- Pereira, M. D. L., Anjos, F. A. dos, & Añaña, E. da S. (2017). SEGMENTACIÓN POST-HOC DEL MERCADO TURÍSTICO EN LOS DESTINOS DEL LITORAL BRASILEÑO. Análisis según los valores personales de los turistas, 26, 662–677.
- Perú, E. C. C. D. del. CONSTITUCIÓN POLÍTICA DEL PERÚ, Pub. L. No. LEY No 29733. Lima, Perú. Retrieved from <http://www4.congreso.gob.pe/ntley/Imagenes/Constitu/Cons1993.pdf>
- Ramírez, D. E., Otero, M. C., & Giraldo, W. (2014). Comportamiento del Consumidor Turístico. *Criterio Libre*, 166–183.
- Rivasplata, J. C., & Sahuanga, C. F. (2012). APLICACIÓN DEL MODELO DE CLUSTERIZACIÓN BASADO EN EL ALGORITMO DE K-MEANS PARA LA SEGMENTACIÓN DE LA MORBILIDAD MATERNA EN EL HOSPITAL SAN BARTOLOMÉ DE LA CIUDAD DE LIMA-2012, 15–21.
- Van, G. (2013). *El tutorial de Python*.
- White, E. G. (2009). *La Educación*. Copyright © 2012, 278. <https://doi.org/10.1016/j.jclepro.2017.04.069>

ANEXOS

ANEXO A. Firma de Emprendedores del Curso Para Recolección de Datos

FACULTAD		: INGENIERIA Y ARQUITECTURA					
EP		: INGENIERIA DE SISTEMAS					
Título del Proyecto/Programa		: TICS PARA TURISMO RURAL COMUNITARIO				Fecha: 16/09/2018	
AÑO		: 2018					
CICLO		: 2018-2					
Nº	Apellidos y Nombres	DNI	Edad	Género		Firma	
				M	F		
1	Cahui Parillo Juana	43158427	33		X	 9623236	
2	Nicolás Mamani Yucra	01247614	64	X		 950930136	
3	Mamani Calcin Saúl Ismael	70159485	20	✓		 950556237	
4	Segundina Calsu Juel	01247870	52			 951008670	
5	David Juli Yaminas	01245234	45	X		 736677656	
6	Roberto Jacansia Jacansia	01292160	40	X		 958299124	
7							
8							
9							
10							
11							
12							
13							
14							

ANEXO B. Anexo B. Participación de la capacitación a emprendedores y recolección de datos de los mismos para aplicar encuesta a los turistas.



ANEXO C. Cuaderno Jupyter de Limpieza de Datos

```
APLICANDO+K-MEANS++PARA+EL+PERFIL+DEL+TURISTA (1).pdf - Adobe Acrobat Reader DC
Archivo Edición Ver Ventana Ayuda
Inicio Herramientas APLICANDO+K-M... x 1 / 29 100% Compartir

In [238]: %matplotlib inline
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt

In [239]: data = pd.read_csv("dataset/Vacacion.csv", low_memory=False, skipinitial...

In [240]: dataset = pd.DataFrame(data)

In [241]: dataset.head()

Out[241]:
```

	LUGAR	P02_1	P02_2	P02_3	P02_4	P02_5	P02_6	P02_7	P02_8	P02_9
0	1	1.0	5.0	6.0	NaN	NaN	NaN	NaN	NaN	NaN
1	1	1.0	5.0	6.0	32.0	75.0	NaN	NaN	NaN	NaN
2	1	1.0	5.0	6.0	75.0	NaN	NaN	NaN	NaN	NaN
3	1	1.0	5.0	6.0	28.0	75.0	NaN	NaN	NaN	NaN
4	1	1.0	5.0	6.0	7.0	NaN	NaN	NaN	NaN	NaN

	...	P65_2	P65_3	P65_4	P65_5	P65_HijosDep	P66	P68	P69
0	...	NaN	NaN	NaN	NaN	2	9	4	99
1	...	NaN	NaN	NaN	NaN	1	3	4	5
2	...	NaN	NaN	NaN	NaN	1	3	6	7
3	...	NaN	NaN	NaN	NaN	2	4	2	99
4	...	NaN	NaN	NaN	NaN	2	2	6	8

	GastoPorPersona	GPC_SA_Ran
0	1201.787274	3
1	1250.000000	3
2	1146.659417	3

```

GastoPorPersona  GPC_SA_Ran
0                1201.787274    3
1                1250.000000    3
2                1146.659417    3
```

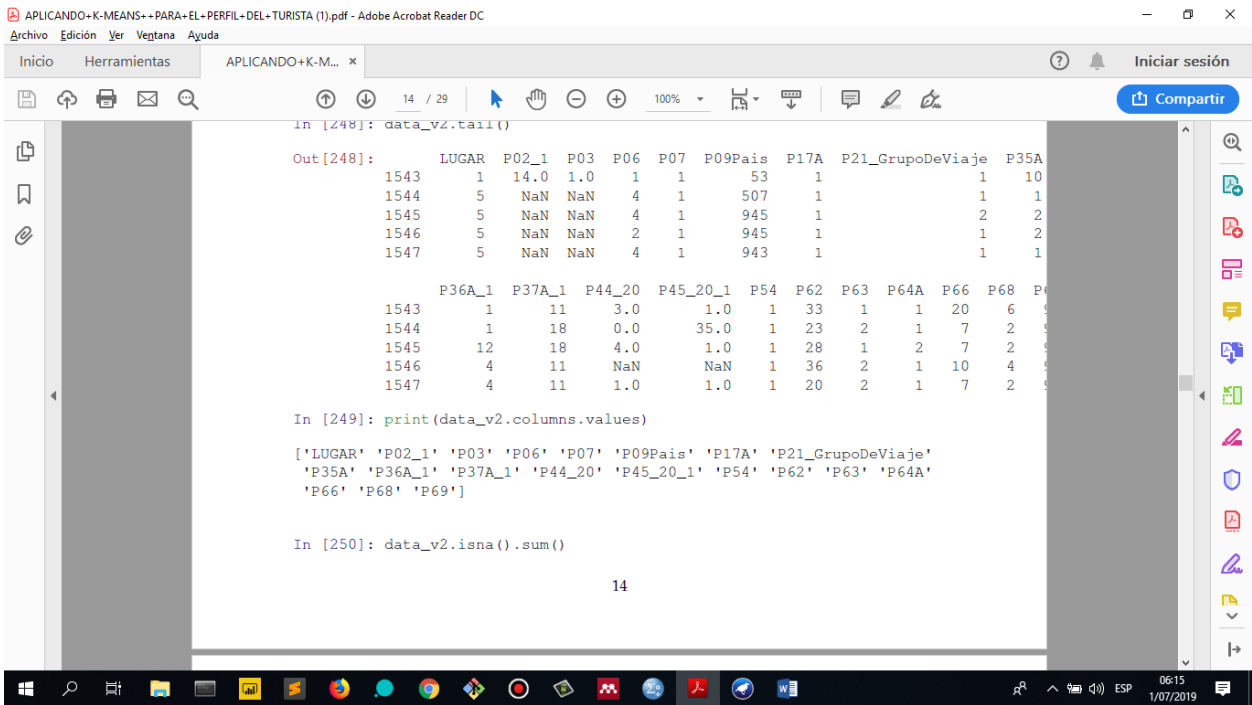
```
APLICANDO+K-MEANS++PARA+EL+PERFIL+DEL+TURISTA (1).pdf - Adobe Acrobat Reader DC
Archivo Edición Ver Ventana Ayuda
Inicio Herramientas APLICANDO+K-M... x 2 / 29 100% Compartir

0.1 Clean the data

In [243]: dataset.isna().sum()
#Obteniendo el numero total de valores faltantes
#data_prueba.isna().sum()

Out[243]:
```

LUGAR	0
P02_1	181
P02_2	408
P02_3	625
P02_4	914
P02_5	1081
P02_6	1190
P02_7	1286
P02_8	1356
P02_9	1401
P02_10	1433
P02_11	1460
P02_12	1481
P02_13	1500
P02_14	1511
P02_15	1522
P02_16	1533
P02_17	1539
P02_18	1543
P02_19	1544
P02_20	1546
P03	181
P06	0



ANEXO D. Encuesta de PromPerú

Archivo Herramientas Vista Cuestionario Vacacionista Extranjero que visita el Peru (PROMPERU 2015).pdf (Solo lectura) - Word (Error de activación de productos)

ESTUDIO DE TURISMO RECEPTIVO – 2015

1 Lima - Extranjeros 2 Puno - Kasani 3 Tacna - Sta. Rosa 4 Lima - Peruanos 5 Tumbes - Aguas Verdes

Lugar de Salida:

Muy buenos días / tardes / noches, mi nombre es... (E: MOSTRAR CARNET DE IDENTIFICACION), y por encargo de PROMPERU estamos llevando a cabo encuestas a los visitantes extranjeros con la finalidad de conocer su opinión respecto al viaje que han realizado a nuestro país, el Perú. Nos gustaría conocer sus opiniones, mientras usted espera para abordar su avión / autobús / transporte.

E: Asegúrese antes de aplicar esta encuesta que el visitante a entrevistar ingresó al Perú y no es un viajero que se encuentra en tránsito en el aeropuerto hacia otro país.

PROGRAMMING: REALIZAR PREGUNTAS 2 Y 3 SOLO A LOS ENTREVISTADOS EN LIMA Y PUNO QUE VINIERON POR VACACIONES O POR COMPRAS PERSONALES COD. 1 O COD. 33 EN P1, CASO CONTRARIO PASE A LA P4.

2. ¿Qué aspectos fueron los que más lo motivaron a visitar el Perú en esta viaje? ¿Algún otro? (E: PROFUNDIZAR. RESPUESTA – ESPONTÁNEA MÚLTIPLE).

3. De todos los aspectos mencionados ¿cuál diría usted que fue el motivo más importante? (PROGRAMMING: MENCIONAR LOS ASPECTOS INDICADOS EN P2. RESPUESTA ÚNICA).

Actividades	P2	P3
Visitar / conocer Machu Picchu	1	1
Visitar sitios o restos arqueológicos (diferentes a Machu Picchu)	2	2
Asistir a espectáculos culturales (teatros, conciertos, exposiciones de arte, etc.)	3	3
Visitar comunidades nativas / campesinas	4	4
Conocer / visitar la ciudad de Lima	5	5
Conocer / visitar la ciudad del Cusco	6	6
Probar la comida peruana / platos típicos	7	7
Participar en festividades patronales o religiosas de los pueblos que visito	8	8
Visitar / conocer áreas / reservas naturales protegidas	9	9
Observar aves	10	10
Observar insectos	11	11

6. ¿Por dónde ingresó al Perú? (RESPUESTA ÚNICA)

1 Lima (Aerop. J. Chavez) 4 Puno (Desaguadero) 94 Otro (ESP.)
 2 Tacna (Sta. Rosa) 6 Tumbes (Aguas Verdes)
 3 Puno (Kasani)

7. ¿Es ésta la primera vez que usted visita el Perú, o lo ha visitado más de una vez? (RESPUESTA ÚNICA)

1 Primera vez
 2 Más de una vez → 7 A. (SOLO PARA QUIENES HAN VISITADO EL PERU MAS DE UNA VEZ. COD. 2 EN P7) Sin considerar esta visita, ¿cuántas veces ha visitado el Perú?
 1. 1 vez
 2. 2 veces
 3. 3 a 5 veces
 4. Más de 5 veces

9. ¿En qué país reside usted? (RESPUESTA ÚNICA)

10 Estados Unidos	6 Inglaterra	11 Argentina	20 Australia
2 Canadá	7 Italia	12 Brasil	21 Japón
3 Alemania	8 Bélgica	13 Bolivia	22 Colombia
4 España	9 Holanda	14 Chile	23 México
5 Francia	10 Suiza	15 Ecuador	24 Otro (ESP.)

Pantallas 1-2 de 17

10. ¿A qué Región /Estado / Provincia pertenece su ciudad de residencia? (LLENADO POR EL ENTREVISTADO)

11. ¿En qué ciudad reside usted? (LLENADO POR EL ENTREVISTADO)

(PROGRAMMING: HACER ESTA PREGUNTA SOLO A LOS VIAJEROS QUE PERMANECIERON EN EL PERU 1 NOCHE O MAS – PREGUNTA 4; EN CASO CONTRARIO IR A LA PREGUNTA 17A)

14. ¿Durante este viaje está visitando Sólo el Perú o también Otros Países? (RESPUESTA UNICA)

1 Solo el Peru → (PROGRAMMING PASAR A LA PREGUNTA 17) 2 Peru y otros países → (PROGRAMMING PASAR A LA PREGUNTA 15)

15. (PROGRAMMING: SOLO PARA QUIENES MARCARON COD. 2 EN P14) Durante este viaje, ¿qué otros países ha visitado o visitará? (PROGRAMMING: ESPONTANEA – RESPUESTA MULTIPLE)

11 Argentina	15 Ecuador	147 Paraguay
12 Brasil	16 USA	150 Uruguay
13 Bolivia	49 Colombia	155 Venezuela
14 Chile	127 Mexico	84 Otros(ESP.)

PROGRAMMING: PARA TODOS LOS ENTREVISTADOS

17A. Según la siguiente tarjeta, (E: MOSTRAR TARJETA P17A) ¿Cómo organizó su viaje? (PROGRAMMING: UNA SOLA RESPUESTA)

1	Contrate todos los servicios por separado
2	Contrate todos los servicios y actividades en un paquete
3	Contrate algunos de los servicios y/o actividades en un paquete y algunos por separado

18.(PROGRAMMING PARA LOS QUE CONTRATARON LOS TODOS LOS SERVICIOS O ALGUNOS SERVICIOS EN PAQUETE – COD 2 O 3 EN P17A) Independientemente de los servicios contratados en su paquete turístico ¿Contrató Ud. algún tour y/o excursión turística durante su estadía en el Perú a una agencia de viajes local?

1. SI 2. NO

19. (PROGRAMMING SOLO PARA LOS QUE VIAJARON CONTRATANDO TODOS LOS SERVICIOS POR SEPARADO – COD 1 EN P17A) ¿Contrató Ud. algún tour y/o excursión turística a una agencia de viajes local durante su estadía en el Perú? (RESPUESTA UNICA)

1. SI 2. NO

21A. ¿Con qué personas realizó su viaje al Perú? (E: LEER RESPUESTAS) (RESPUESTA MULTIPLE)

1	SOLU (PROGRAMMING: PASE A PREGUNTA 25) ALIAR/PANAJUE...
2	Con mi pareja
3	Con mis hijos
4	Con mis nietos
5	Con mis padres
6	Con amigos o parientes, sin niños
7	Con amigos o parientes, con niños
8	Con compañeros de trabajo
9	Otros (ESP.)

23. En los gastos que va a declarar usted, ¿Cuántas personas están consideradas, incluido/a usted?

N° de personas:

PROGRAMMING: VERIFIQUE RESPUESTA SI EL ENTREVISTADO VINO SOLO

PROGRAMMING (HACER LAS SIGUIENTES PREGUNTAS SOLO A QUIENES LOS QUE CONTRATARON TODOS LOS SERVICIOS O ALGUNOS SERVICIOS EN PAQUETE COD 2 O 3 EN P17A, DE LO CONTRARIO PASAR A LA PREGUNTA 29)

24. Considerando todos los países comprendidos en su viaje y a todas las personas que dependen de usted en este viaje ¿Cuál ha sido el costo del paquete turístico que adquirió? _____

25. Específicamente en lo que se refiere a su viaje al Perú, según la siguiente tarjeta, (E: MOSTRAR TARJETA P25 / P27/ P29) ¿Qué servicios incluyó el paquete turístico que Ud. adquirió? (MULTIPLE)

RUBROS	Preg. 25 Servicios incluidos en el paquete turístico (PROGRAMMING: RESPUESTA MULTIPLE)
Transporte aéreo internacional (entrada y/o salida del Perú)	1
Transporte terrestre internacional (entrada y/o salida del Perú)	2
Transporte aéreo dentro del Perú entre ciudades	3
Transporte terrestre dentro del Perú entre ciudades	4
Traslados internos en las ciudades visitadas (alquiler de autos, buses, taxis)	5
Alojamiento	6
Restaurantes	7
Alimentos, bebidas y tabaco	8
Centros nocturnos y discotecas	9
Museos y exposiciones	10
Actividades deportivas, taumias, artísticas	11
Visitas a atractivos turísticos (centros arqueológicos, áreas naturales, etc.)	12
Otros (ESP.)	W4
Otros (ESP.)	W4
Otros (ESP.)	W4

26A. ¿Cómo adquirió su paquete? (LEER OPCIONES DE RESPUESTA)

- En una agencia de viajes física
- A través de internet → 26B. (SÓLO PARA QUIENES ADQUIRIERON SU PAQUETE A TRAVÉS DE INTERNET, COD. 2 EN P26A)
- No lo sé/ ¿A través de qué página? _____
- La empresa lo compró por mí

27. Según esta misma tarjeta (E: SIGA MOSTRANDO TARJETA P25 / P27/ P29) ¿En qué rubros realizó gastos adicionales a lo incluido en su paquete? (MULTIPLE)

28. ¿Cuánto fue lo que gastó en...? (PROGRAMMING: PREGUNTAR POR CADA RUBRO MENCIONADO EN P.27. SOLICITAR LOS MONTOS, SUMARLOS Y CONFIRMAR CON EL ENTREVISTADO QUE EL RESULTADO ES EFECTIVAMENTE EL MONTO GASTADO. LUEGO PASAR A P31A).

Cuadro de servicios RUBROS	Preg. 27	Para la Preg. 28	
	Rubros no incluidos en el paquete turístico (PROGRAMMING: RESPUESTA MULTIPLE)	Moneda: Dólares Americanos (Montos)	(%)
Transporte aéreo internacional (entrada y/o salida del Perú)	1		
Transporte terrestre internacional (entrada y/o salida del Perú)	2		
Transporte aéreo dentro del Perú entre ciudades	3		
Transporte terrestre dentro del Perú entre ciudades	4		
Traslados internos en las ciudades visitadas (alquiler de autos, buses, taxis)	5		
Alojamiento	6		
Restaurantes	7		
Alimentos, bebidas y tabaco	8		
Centros nocturnos y discotecas	9		
Museos y exposiciones	10		
Actividades deportivas, taumias, artísticas	11		
Visitas a atractivos turísticos (centros arqueológicos, áreas naturales, etc.)	12		
Artesanías / souvenirs	13		
Prendas de vestir y calzado	14		
Otros (ESP.)	W4		
Otros (ESP.)	W4		
Otros (ESP.)	W4		
No realizó gastos adicionales a mi paquete (PROGRAMMING: IR A P29)	W0		
TOTAL:			

E: Si existe alguna observación o comentario para algunas de las preguntas: P25 / P26 / P27,

indicarlo a continuación (E: INDICAR PREGUNTA Y RUBRO):

Observación 1

ANEXO E. MAPIC

VARIABLE FÁCTICA	DIMENSIONES	INDICADORES
Identificación del perfil del turista que realiza Turismo Rural Comunitario	Motivaciones	Segmentos de perfil de turistas.
	Sociodemográficas	
	Comportamiento	
	Sostenibilidad	
VARIABLE TEMÁTICA	EJES TEMÁTICOS	SUB EJES TEMÁTICOS
Desarrollo de software	- Metodología de desarrollo de aplicaciones).	- Persistencia de datos
	- Desarrollo de app web - Desarrollo de app Móvil	- Offline-first - Aplicaciones nativas e híbridas. - Desarrollo de Aplicaciones nativas con herramientas web. - Python -JavaScript y frameworks js.
Minería de datos	Minería de datos	Algoritmos de minería de datos.
VARIABLE PROPOSITIVA	EJES PROPOSITIVOS	SUB EJES PROPOSITIVOS
Aplicación del algoritmo k-means	Aplicación del algoritmo k-means.	Limpieza de datos
		Pruebas y validación del algoritmo con datos reales.