

TRABAJO DE GRADO
Opción Seminario-Diplomado.

Análisis de Datos sobre Dispositivos Móviles

Corporación Universitaria Remington
Facultad de Ingeniería
Ingeniería de Sistemas

¹Juan Carlos Betancourt Gutiérrez
²Ivonne Castaño Osorio
³John Edison Amortegui
Seminario
2024

¹ Estudiante de Noveno Semestre de Ingeniería de Sistemas sede Uniremington Pereira. E-mail: juan.betancourt.6645@miremington.edu.co

² Tutor temático Seminario en Big Data sede Pereira: ivonne.castano@uniremington.edu.co

³ Tutor metodológico Seminario en Big Data sede Pereira Email: john.amortegui@uniremington.edu.co

Dedicatoria

Dedico este trabajo a mi familia quienes siempre han creado en mí, dándome ejemplo de superación, humildad y sacrificio, donde me enseñaron a valorar cada logro que he conseguido hasta el momento, también quiero dedicarle este trabajo a los docentes que un momento me enseñaron un poco de su conocimiento para aplicarlos en mi vida profesional.

Tabla de contenidos

Resumen.....	4
Pregunta Orientadora de la búsqueda	6
Metodología de búsqueda de la información.....	7
Gráfico Pareto	9
Gráfico de barras.....	9
Gráfico de áreas	10
Sustentación teórica de la pregunta.....	11
Modelado	11
Estadística descriptiva.....	11
Coeficiente de correlación	12
Covarianza	12
Regresión Estadística	12
Regresión Análisis de Varianza.....	13
Cantidad de celulares vendidos en el mundo en el año 2023	¡Error! Marcador no definido.
Conclusiones.....	17
Referencias.....	18

Resumen

Cuando se habla de ciencia de datos, lo podemos definir como métodos científicos, procesos, algoritmos donde se puede obtener información sobre determinados datos y sacar conclusiones verídicas para realizar cualquier análisis o sacar conclusiones. Hoy en día y con el crecimiento global de la tecnología se puede determinar que la ciencia de datos se vive diariamente con tan solo un Click desde nuestros dispositivos móviles o computadoras, por ejemplo, en Colombia se realiza según Cajamarca (2024) un promedio 35 millones de transacciones a través de plataformas bancarias, la ciencia de datos también se puede generar cuando hacemos uso de nuestras redes sociales. En este orden de ideas Ortega (2024) afirma que la ciencia de datos es importante para generar ideas y tomar decisiones a nivel empresarial, optimizar procesos, innovación y competitividad, conocimiento científico, detección de fraudes, gestión de recursos y planeación estratégica. Uno de los procesos más importantes de la ciencia de datos que conlleva analizar y comprender datos, sacar conclusiones y resolver problemas son: obtener la información, donde se puede concluir el tipo de datos que se tienen en una base, limpiar los datos. Es importante realizar depuraciones y trabajar con datos que nos ayuden a interpretar mejor el mercado, analizar la colocación de los datos para poder analizarlos y encontrar algún patrón inusual. Aprendizaje de datos, en este caso un ingeniero de datos o un científico de datos, escribe algunas instrucciones para que un algoritmo las estudie y pueda llegar a un resultado determinado por último se tiene la manipulación de datos, en este caso se le muestra los resultados a una organización determinada y así medir efectividades. La metodología de la ciencia de datos tiene como objetivo responder preguntas como, cual es el problema que se está tratando de resolver,

como usar los datos para resolver la pregunta, que datos se necesitan para responder dicha pregunta, de donde provienen los datos, los datos son representativos del problema, que trabajo adicional se requiere para manipular la información, de qué manera se puede visualizar los datos para obtener la respuesta, el modelo utilizado responde a la pregunta, poner el modelo en práctica, obtener comentarios constructivos para resolver la pregunta. (TuDashboard, 2021)

Palabras Claves: Teléfonos móviles, Smartphone, Android, iOS, Versión Sistema Operativo, Almacenamiento Interno.

Pregunta orientadora de la búsqueda

En la actualidad se puede observar que la gran mayoría de las personas se fijan en determinadas características de un dispositivo móvil, las cuales pueden influir en el precio de estos, sin embargo, estas características no deberían ser las únicas en las que una persona se fije, si no, abordar más en el tema de lo que tiene el dispositivo como tal y analizar cuál en realidad será su uso final.

Por eso se hizo un análisis de las principales características de los dispositivos más importantes del mercado, para que las personas puedan hacer una comparación más detallada de lo que en realidad necesitan, si es para uso personal o laboral. Por lo anterior, es importante formular la siguiente pregunta: ¿Se puede realizar un análisis de datos con las características de los celulares que determinen su precio en el mercado?

Metodología de búsqueda de la información.

La metodología CRISP-DM tiene como objetivo 10 preguntas básicas, es una secuencia prescrita, del problema al enfoque, trabajando con los datos, derivando respuestas, algunos de los aspectos claves de esta metodología son identificación de fuentes de datos, evaluación y calidad de los datos, exploración de los datos, acuerdo sobre el formato de los datos, recopilación y selección de datos y limpieza y transformación de los datos.

En este caso el enfoque analítico sobre las múltiples marcas de celulares que hay en el mercado es como llegar al consumidor, muchas de estas personas se basan en analizar cada dispositivo móvil y elegir el que más le supla las necesidades diarias. Los requisitos de datos los podemos interpretar de la siguiente manera, realizando breves preguntas o encuestas donde se van a recopilar distintas respuestas y así realizar un breve análisis sobre los diferentes gustos de la gente.

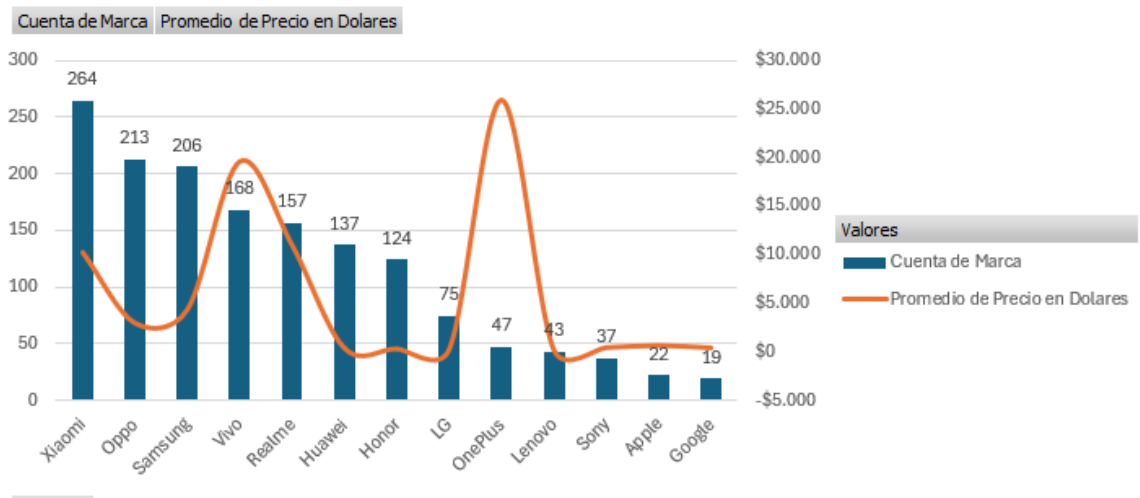
Para la recopilación de datos se realiza un sondeo sobre una población determinada exponiendo cada característica de un dispositivo móvil como, marca, sistema operativo, resolución, durabilidad de la batería, memoria RAM, almacenamiento, calidad de la cámara, entre otras.

A continuación, una breve descripción de algunas marcas que se encuentran en el mercado, en este caso se tiene una base de datos con distintas marcas y dispositivos móviles la cual contiene sus características más relevantes como, sistema operativo, memoria RAM, tipo de batería, precio Etc.

Para llegar a este análisis se realizó un promedio donde se puede determinar el precio versus la cantidad de dispositivos vendidos, se determinó el promedio donde 1 es verdadero y 0 es falso, con estos valores se procede a sumarlos para determinar la cantidad de características para llegar a un valor determinado que muestre cuál de las marcas es mejor en este ámbito y por último se determinó la capacidad de almacenamiento de cada marca.

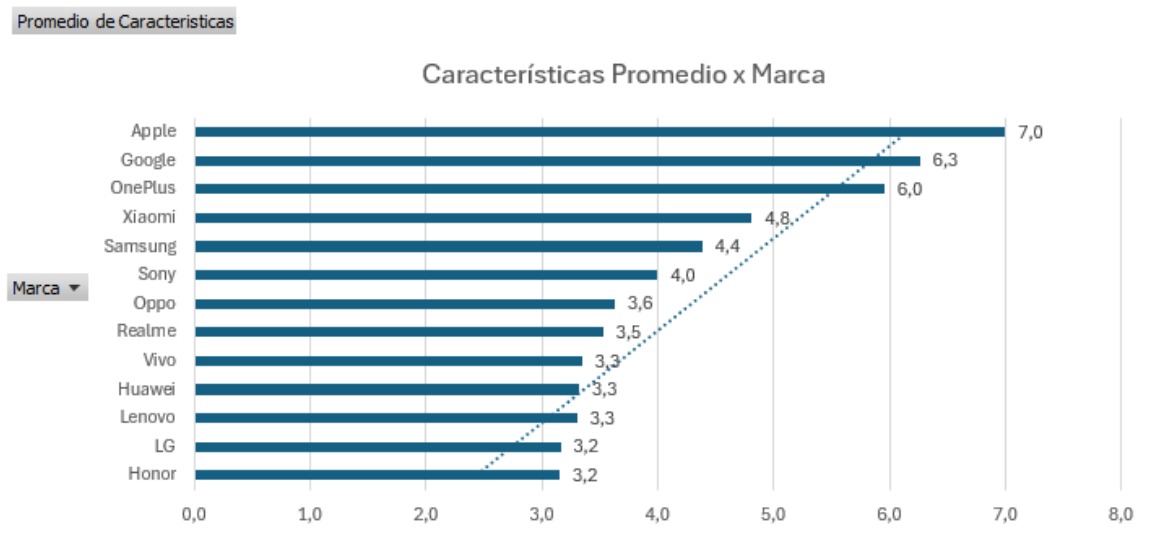
En la metodología, en el enfoque del problema se puede determinar que el precio de cada celular va depender del almacenamiento entre mayor capacidad el valor aumenta, en el análisis con los datos se pudo determinar que el consumidor final es el que determina la capacidad de recursos que necesita para su celular independientemente del precio, en conclusión en el mercado actual se puede encontrar diferentes modelos, estilos, colores, materiales, pero la principal causa de la variación del precio en el mercado es su almacenamiento, hoy en día con la tendencia que va en crecimiento del *Streaming*, las redes sociales los usuarios finales prefieren celulares con mayor almacenamiento para tener eficiencia a la hora de guardar datos. La cantidad de equipos vendidos no siempre es proporcional a las características y su valor en dinero.

Gráfico 1. Pareto.



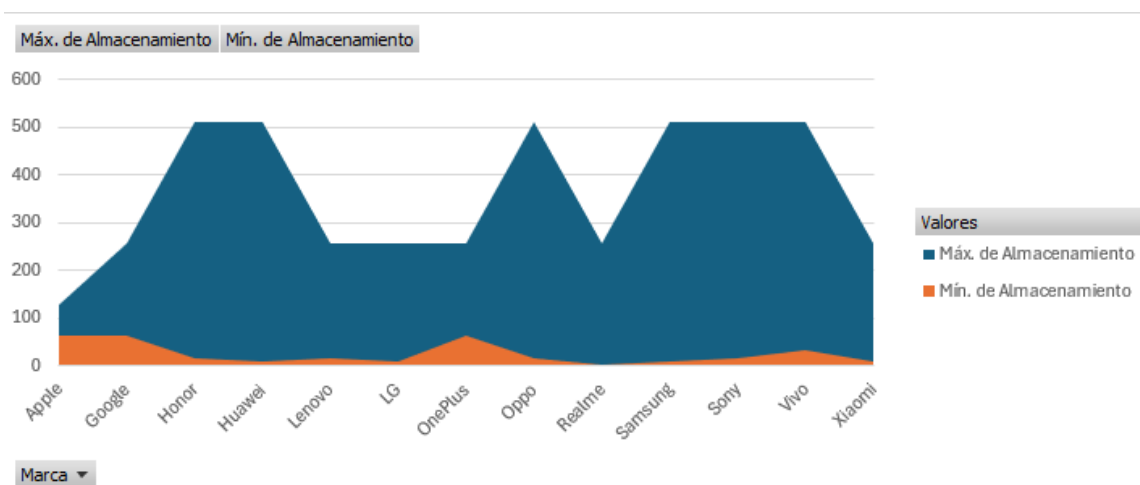
Como se puede observar en el gráfico (Ver Gráfico1. Pareto) la marca Apple sobresale del promedio de características de las demás marcas que existen con relación a la base de datos. El promedio de ventas de la marca HONOR es superior a las de SONY.

Gráfico 2. De barras.



En el gráfico (Gráfico 2. De barras) se puede observar el promedio sobre el almacenamiento de los dispositivos móviles.

Gráfico 3. De Áreas.



En el gráfico (Gráfico 3. De Áreas) se puede observar el promedio de almacenamiento de los dispositivos móviles.

Sustentación teórica de la pregunta

Modelado

Tabla 1. Estadística descriptiva.

Almacenamiento	Precio
Media: 109,1646825	Media: 13393,6905
Error Típico: 1,914299918	Error Típico: 1139,805298
Mediana: 128	Mediana: 3000
Moda: 128	Moda: 2000
Desviación Estándar: 74,43648434	Desviación Estándar: 44320,69548
Varianza de la Muestra: 5540,790201	Varianza de la muestra: 1964324048
Curtosis: 6,27946667	Curtosis: 68,16326077
Coefficiente de asimetría: 1,842679732	Coefficiente de asimetría: 7,351993439
Rango: 511	Rango: 659588
Mínimo: 1	Mínimo: 400
Máximo: 512	Máximo: 659988
Suma: 165057	Suma: 20251530
Cuenta: 1512	Cuenta: 1512
Nivel de confianza: (95,0%) 3,754966514	Nivel de confianza: (95,0%) 2235,768238

En la estadística descriptiva se determinaron dos variables que son el almacenamiento y el precio, donde podemos determinar lo siguiente:

El valor promedio de almacenamiento de los celulares es de 109 GB en adelante, la mediana es el valor que se encuentra justo en el medio cuando los datos de orden de menor a mayor en este caso son de 128 GB por su promedio en valores, realizando un análisis sobre el almacenamiento, podemos concluir que los celulares más vendidos son lo que tienen 128 GB de almacenamiento.

Tabla 2. Coeficiente de correlación.

	<i>Memoria RAM</i>	<i>almacenamiento</i>	<i>precio</i>
memoria RAM	1		
almacenamiento	0,699045291	1	
precio	0,08650728	0,079650598	1

La correlación es positiva dado que ningún valor es negativo en el almacenamiento y su promedio está por encima de 0,5, en el precio la correlación es negativa dado que está en 0,0 por su variación de valor.

Tabla 3. Covarianza.

	<i>Memoria RAM</i>	<i>almacenamiento</i>	<i>precio</i>
memoria RAM	7,292914182		
almacenamiento	140,4746814	5537,125657	
precio	10350,61402	262599,6459	1963024892

El análisis realizado en la covarianza se refleja en el precio, entre más grande sea el promedio tienen mayor relación, teniendo en cuenta que el precio varía dependiendo del almacenamiento.

Tabla 4. Regresión Estadística.

Coeficiente de correlación múltiple	0,811067148
Coeficiente de determinación R ²	0,657829919
R ² ajustado	0,655821887
Error típico	955,3308653
Observaciones	499

Tabla 5. Regresión Análisis de Varianza.

	<i>Grados de libertad</i>	<i>Suma de cuadrados</i>	<i>Promedio de los cuadrados</i>	<i>F</i>	<i>Valor crítico de F</i>
Regresión	1	873792980	873792980	957,4165545	5,7581E-118
Residuos	498	454503217	912657,0622		
Total	499	1328296197			

	<i>Coefficientes</i>	<i>Error típico</i>	<i>Estadístico t</i>	<i>Probabilidad</i>	<i>Inferior 95%</i>	<i>Superior 95%</i>	<i>Inferior 95,0%</i>	<i>Superior 95,0%</i>
Intercepción	0	#N/D	#N/D	#N/D	#N/D	#N/D	#N/D	#N/D
almacenamiento	13,2500225	0,428219214	30,94214851	4,6759E-118	12,40868352	14,09136149	12,40868352	14,09136149

Realizando un análisis sobre el mercado de los celulares podemos determinar que en la actualidad el mercado es muy dinámico y competitivo, por la rápida implementación de nuevas tecnologías y la demanda que va en crecimiento a nivel mundial. Una de las tendencias hoy en día en la red 5G que algunos fabricantes ya la tienen implementada en algunos celulares esto ofrece mayor velocidad en conexión. En el análisis también se puede determinar que las marcas top en el mercado son Apple y Samsung siendo las primeras en ventas a nivel mundial, pero marcas como Oppo, Xiaomi, Huawei, vienen cogiendo fuerza en diseño, cámara y almacenamiento. Este análisis fue enfocado en el almacenamiento de los celulares, donde algunos modelos de entrada solo ofrecen entre 32 GB o 64 GB de almacenamiento, estos suelen ser un poco más baratos a comparación con los que tienen un almacenamiento mayor de 128 a 512

GB. Algunas estrategias de los fabricantes de celulares es lanzar al mercado diferentes versiones de un mismo modelo con la diferencia de la capacidad de almacenamiento y así ofrecer precios competitivos con las necesidades de los usuarios finales. Tenemos otros factores que pueden determinar el precio de un celular como el procesador, este determina la eficiencia y el rendimiento del celular, memoria RAM, es la que se encarga de ejecutar varias aplicaciones sin problemas, el tipo de pantalla si es OLED o AMOLED, cámara, acá se tienen características avanzadas como grabar videos en modo cine, estabilización y zoom óptico, estas características y muchas más pueden determinar el precio de un celular.

Algunas observaciones sobre las diferentes marcas que tenemos en la base de datos; Apple por su precio promedio vende menos unidades que el promedio vendió por Xiaomi, mientras que la marca Apple si sitúa de primero en la lista con su promedio más alto en características en los equipos. Samsung es la marca que por su precio promedio es más asequible que la marca Lenovo, donde sus características son más sobresalientes que la marca Lenovo, y el almacenamiento de Samsung es mayor al de Lenovo, el promedio más alto de almacenamiento lo tienen las marcas con sistema operativo Android, aunque el promedio de características de la marca OnePlus este entre una de las mejores, su almacenamiento no es muy competitivo con otras marcas del mercado.

Tabla 6. Número de ventas por según marca 2023.

<i>Nombre de Celular</i>	<i>Cantidad Vendidos</i>
iPhone 14 pro-Max	26,5 millones
iPhone 14 pro	21 millones
iPhone 14	16,5 millones
Galaxy A14	15,5 millones
iPhone 13	12,4 millones
Galaxy S23 ultra	9,6 millones
Galaxy A14 5G	9 millones
Galaxy A54 5G	8.8 millones
Galaxy A34 5G	7,1 millones
IPhone 11	6.9 millones

Fuente: (Valora Analitik, 2024)

Para complementar la tabla, a continuación, un breve resumen de los dispositivos móviles más vendidos en Colombia en el año 2023 según cada empresa:

Según telefónica movistar estos fueron los celulares más vendidos: OppoA57 de 128 GB, Samsung A22 5G de 128 GB, Xiaomi Note 12 de 128 GB, Honor X7A de 128 GB, Xiaomi Note 11 de 128 GB, Moto G5 de 128 GB. Por parte de la compañía Wom estos fueron los más vendidos: Xiaomi Redmi Note 12 de 128 GB, Motorola G22de 128 GB, Xiaomi Redmi Note 11 128 GB, Motorola G33 128 GB, Motorola E22l 64 GB. La compañía claro no da muchos detalles de sus ventas por políticas de seguridad, esas fueron sus marcas más vendidas: Samsung, Motorola, Xiaomi, Oppo, Honor. Tigo informa que los colombianos siguen prefiriendo el sistema operativo Android y su capacidad se convirtió en la estrategia clave de ventas, estos fueron sus dispositivos más

vendidos: Xiaomi Redmi Note 11 4G 128 GB, Xiaomi Redmi Note 12 4G 128 GB,
Motorola G22 128 GB, Motorola E13 64 GB, Samsung A24 128 GB. (Hernandez, 2023).

Conclusiones

Es importante conocer las necesidades de las organizaciones para saber cómo tratar los datos y así contribuir con las metas y objetivos. Las grandes empresas de dispositivos móviles están llegando al punto de que la población sea más dependiente de ellos. El cifrado de los datos es importante para evitar cualquier tipo de ataque o hackeo por parte de personas inescrupulosas. La ciencia de datos se utiliza para el análisis descriptivo, análisis de diagnóstico, análisis predictivo, análisis prescriptivo.

La ciencia de datos tiene beneficios importantes para las empresas como: descubrir patrones desconocidos, innovar en nuevos productos y soluciones, optimizar procesos en tiempo real. El proceso de la ciencia de datos tiene como objetivo, obtener datos, depurar datos, explorar datos, modelar datos, interpretar resultados. Uno de los factores determinantes a la hora de comprar un celular sea gama media o alta suele depender del almacenamiento interno.

La sociedad de hoy en día puede aprovechar la ciencia de datos y sacar buenos beneficios, para esto es necesario garantizar la privacidad de los datos y asegurar de que se utilicen de manera correcta y así tomar decisiones adecuadas para el día a día.

Lista de referencias

Juan Neira (2024). *Los celulares más vendidos en el año 2023*.

Iván Cajamarca (2021). “Al mes se realizan 35 millones de operaciones” *70 la República*.

Lina Hernández (2023). *Los celulares más vendidos en Colombia El tiempo*.

TuDashboard (2021). Que es la ciencia de datos y cuál es su objetivo.

Edwin Puertas del Castillo *Ciencia de datos, una excelente oportunidad para el perfil profesional*
(2022). *Universidad Tecnológica de Bolívar*. <https://acortar.link/vWowGC>

Ing. Elías Porras Umaña (2021). *Ciencia de datos, la revolución apenas comienza* (Umaña, 2021)