



Tecnológico de Monterrey

Campus Toluca

Primer Avance del Proyecto

Materia: IN3038.1

Laboratorio de Diseño y Optimización de Operaciones

Profesora: M. en C. Ana Luisa Masetto Herrera

Equipo número 3:

Alejandra Velazquez Bastida | A01368039

Arlin San Juan Pezat | A01368226

Yessica Vidal Castellanos | A01366760

Alejandro Gabriel Hernandez | A01367757

Juan Carlos Robles Guicho | A01368398

Fecha de entrega: *Lunes 6 de septiembre del 2021*

Semestre Agosto- Diciembre 2021

1. Etapa 1- Comprensión del Negocio.

1.1 Descripción de la situación actual.

Hisense llega a México en 2011 y en tan solo una década, el país se posicionó en el tercer mercado más importante para la empresa, debido a la relación costo beneficio sus diferentes productos electrodomésticos y de línea blanca. Se sabe que entre el 20% y 25% de la producción se queda para consumo local [1]. Derivado de la pandemia del COVID-19, la escasez de chips y el aumento en los precios del cobre, aluminio y plástico, la cadena de suministro de Hisense se ha visto afectada drásticamente. La empresa ha sufrido una disminución en la participación del mercado de venta de smartphones, cayendo del 0.05% al 0.022% del 2020 a 2021. Aún con todo esto la empresa ha priorizado a México como uno de los principales mercados para sus productos (Statista Research Department, 2021).

Por contexto como el anterior nace la importancia de realizar un proyecto como este, un ingeniero industrial puede tomar decisiones con base en datos analizados estadística y matemáticamente con programación y no solo por intuición. Esto constituye uno de los perfiles profesionales más relevantes en la actualidad; el poder entender y analizar sistemas complejos.

1.2 Entender y describir la problemática. (en términos del negocio)

Es de suma importancia que la empresa comience a generar estrategias que le permitan aprovechar el creciente mercado de celulares, debido a su gran desempeño el año pasado pese a la pandemia del COVID-19. Además, una correcta alineación de la producción con la demanda le permitirá a Hisense afrontar los escenarios presentes. Para este problema un ingeniero industrial con conocimientos en ciencia de datos puede auxiliar a la empresa en el proceso de toma de decisiones.

1.3 Entender y describir la problemática. (en términos de ciencia de datos)

La empresa Hisense requiere la construcción de portafolios de productos (predicción de demanda) para sus diferentes puntos de venta. Para nuestro caso de estudio tenemos un problema de regresión, pues vamos a pronosticar un valor numérico con base en datos **estructurados** que se nos han brindado. Con este proyecto buscamos dar respuesta a la interrogante: **¿Cuántos celulares se venderán en cada punto de venta?** Generando así una reducción en los costos logísticos y un aumento en la satisfacción del cliente.

1.4 Plasmar los objetivos. Nuestro objetivo como equipo es lograr el entendimiento y la correcta aplicación de un proyecto de ciencia de datos, con la comprensión y estructura de la metodología CRISP-DM.

Etapas	Objetivos	Indicador de éxito
Business Understanding	Entender el contexto, historia, objetivos, criterios de éxito, riesgos, recursos, terminología y la problemática de la empresa.	Reporte y desglose de sucesos relevante de la empresa.
	Identificar los objetivos del data mining.	Tabla de objetivos e indicadores por fase.
	Definir plan inicial.	Diagrama de Gantt.
Data Understanding	Recolección/ solicitud de datos.	Base de datos completa.
	Comprensión de la terminología interna de la empresa.	Diccionario de las variables a utilizar.
	Descripción de datos.	División y clasificación de variables.
	Verificar calidad de datos/Identificar el tamaño y tipo de archivo de nuestros datos.	Lectura de archivo CSV, reporte de calidad datos.
Data preparation	Limpieza correcta de datos.	Datos estandarizados, congruentes y en orden.
	Selección del software o plataforma para el manejo y análisis de datos.	Software, plataforma instalada.
	Archivo preparado para el software a utilizar (CSV, XML, HTML, etc.)	Archivo listo.
Modeling	Selección correcta de algoritmo para pronosticar y técnicas a emplear.	Notas y ecuaciones listas.
	Generar test design.	Fracción de datos a utilizar en test design.
	Construcción efectiva del modelo	Código funcional.
	Definición de parámetros del modelo.	Métrica de eficiencia del modelo.
Evaluation	Evaluación efectiva de resultados.	Resultados alineados a criterios de evaluación de la empresa.
	Revisión y mapeo de proceso.	Checklist de proceso.
	Definir next steps.	Planeación de cierre.
Deployment	Definir plan de documentación.	Checklist de etapas del proceso a documentar.
	Definir plan de monitoreo y control.	Criterios / actividades de monitoreo y control definidos.
	Documentación de experiencia.	Reporte final.

Fig 1. Tabla de objetivos por etapa. Ver completo:

<https://docs.google.com/spreadsheets/d/1s1dyiofhoxb-HrmjxUCh9EiTJFuIDLk/edit?usp=sharing&ouid=101688482481224037026&rtpof=true&sd=true>

1.5 Estructurar el proyecto y hacer un plan preliminar.

A lo largo de este documento se desarrollará un proyecto de ciencia de datos enfocado a dar solución al proceso de generación de portafolios de productos, una predicción de la demanda para los diferentes puntos de venta de la empresa en estudio. El proyecto se estructurará tomando como base la metodología CRISP-DM y la aplicación de diversas herramientas de ciencias de datos, para ello se seguirá el calendario de actividades plasmadas en el siguiente diagrama de Gantt: Anexo. Diagrama de Gantt: <https://docs.google.com/spreadsheets/d/1N83SSjAv8x7GcwoVv-xdbqd3mV9uykCN/edit?usp=sharing&ouid=101688482481224037026&rtpof=true&sd=true>

2. Etapa 2- Comprensión de los datos.

2.1 Describir los datos crudos.

La empresa nos proporcionó una base de datos en la cual tenemos información de las unidades vendidas desde el 1ro de junio del 2019 hasta el 31 de marzo del 2020.

Tipo de archivo: **Excel.csv**

Dimensión de los datos: **14 variables con 23032 registros en total.**

No.	Variable	Descripción	Tipo de dato	Valor min	Valor max	Nivel	Ejemplo
1	Punto_de_venta	Puntos de venta donde la empresa vende	Character			1538	"1 poniente", "5 de mayo zmm", "acayucan"
2	Fecha	Fecha de la venta del equipo	Character			301	31/12/19
3	Mes	Mes de la venta del equipo	Character			11	"AGOST", "FEB", "1", "10"
4	Año	Año de la venta del equipo	Number	2020	2019	2	"19", "2019", "2020"
5	Num_ventas	Unidades vendidas del equipo	Number	1	1	1	1
6	Sku	Número de serie del equipo vendido	Character			28	"N.HIF24AZ", "N.HIF24NG", "N.HIL675B"
7	Marca	Marca del equipo vendido	Character			5	"Hisense-Hisense", "hisense"
8	Gamma	Categoría del equipo en el mercado	Character			1	"baja"
9	Costo_promedio	Costo promedio de los equipos vendidos	Number	0	2067	22	"1982.981979", "2067.007825"
10	Zona	Zona donde se vendió el equipo	Character			9	"centro occidente", "centro sur"
11	Estado	Estado donde se vendió el equipo	Character			35	"aguascalientes", "baja california"
12	Ciudad	Ciudad donde se vendió el equipo	Character			210	"zitacuaro", "zumpango"
13	Latitud	Latitud de la ciudad donde se vendió el equipo	Number	14.9	1958275	1515	"21.0302", "21.03042", "21.03051"
14	Longitud	Longitud de la ciudad donde se vendió el equipo	Number	-990229	-86.8	1473	"-99.07311", "-99.06958", "-99.05718"

Fig 2. Diccionario de variables.

2.2 Detectar problemas de calidad.

No.	Variable	Problema de calidad	¿Por qué?
1	Punto_de_venta	Coherencia	En la sucursal de acayucan se encuentra otra sucursal capturada como acayuckan, por lo que se ve incongruencia en la captura del nombre de las sucursales.
2	Fecha		
3	Mes	Coherencia	Se usa diferente notación para la identificación del mes.
4	Año	Coherencia	Se usa diferente notación para la identificación del año.
5	Num_ventas		
6	Sku	Representación	El número de caracteres en la serie no es congruente.
7	Marca	Unicidad	El nombre de la marca no está estandarizado para todos los registros.
8	Gamma		
9	Costo_promedio	Compleitud	Hay valores faltantes.
10	Zona	Coherencia	Se usan mayúsculas para la identificación de la zona del golfo de México.
11	Estado	Precisión	Existen registros de ciudades como estados.
12	Ciudad		
13	Latitud	Representación	El número de decimales no es igual en todos los datos.
14	Longitud	Representación	El número de decimales no es igual en todos los datos.

Fig 3. Tabla de problemas de calidad.

2.3 Actividades a futuro

Conforme a lo detallado en el diagrama de Gantt previamente anexo, como equipo debemos enfocarnos en la preparación de los datos a través de la limpieza, análisis y selección de variables relevantes para avanzar a las siguientes etapas. Con esto en mente, es importante agendar periódicamente sesiones de trabajo donde estemos presentes todos los integrantes del equipo y colaboremos activamente a garantizar datos confiables. De esta forma, tendremos conocimiento del proceso que lleve a acciones directas que apoyen a las estrategias y acciones competitivas para Hisense.

3. Referencias Bibliográficas:

- [1] (n.d.). Sobre Hisense. Se recuperó el septiembre 01, 2021 de <https://www.hisense.es/sobre-hisense/>
- [2] Cristina O. (2021, julio 17). Hisense ve en México una prioridad en ventas y producción. Periódico digital Milenio, Se recuperó el septiembre 2, 2021 de: <https://www.milenio.com/negocios/hisense-ve-mexico-prioridad-ventas-produccion>
- [3] Statista Research Department (2021). *Marcas de smartphones con mayor cuota de mercado en México 2021*. Disponible en: <https://es.statista.com/estadisticas/1077738/participacion-mercado-marcas-smartphones-mexico/>
- [4] Steve, O. (2020). *Con todo y COVID, los mexicanos nunca habían comprado tantos smartphones de entre 16,000 y 20,000 pesos como en 2020*. Xataka México. Disponible en: <https://www.xataka.com.mx/celulares-y-smartphones/todo-covid-mexicanos-nunca-habian-comprado-smartphones-16-000-20-000-pesos-como-2020>