

Propuesta de Proyecto para Introducción a la Ciencia de Datos

Andrés Vargas, Danny Ucho & Juan Nebel – Facts & Stats

Escuela Superior Politécnica del Litoral

Propuesta de Proyecto para Introducción a la Ciencia de Datos

En este reporte se detallan dos propuestas de proyectos tras haber realizado una investigación por parte de los integrantes del grupo llamado “Facts & Stats” conformado por Andrés Vargas, Danny Ucho y Juan Nebel.

Análisis transaccional

En base a registros de compra se propone uno de los siguientes puntos:

1. Realizar un perfilamiento de los compradores en función a los artículos que éste compra.
2. CLV (Customer Lifetime Value analysis)
3. RFM (Recency, Frequency and Monetary)

Área propuesta

Sector económico – Comercialización de productos

Métodos de Recolección

1. Data sets disponibles en Kaggle
2. Tentativamente datos proveídos de Tía S.A.

Descripción del Dataset

El conjunto de datos anónimos incluye 64.682 transacciones de 5.242 SKU vendidos a 22.625 clientes durante un año.

1. Date of Sales Transaction
2. Customer ID
3. Transaction ID
4. SKU Category ID

5. SKU ID
6. Quantity Sold
7. Sales Amount (Unit price times quantity. For unit price, please divide Sales Amount by Quantity.)

El campo “Date of Sales Transaction” es particularmente importante útil para definir intervalos temporales en caso de ser necesario. Por otra parte el SKU ID y el SKU Category permite definir el producto comprado; esto da paso hacer análisis de preferencia de compra por cliente gracias a la asociación con el campo “Customer ID”.

Preguntas de negocio

1. **¿Cómo se verían los segmentos de los clientes basados en el *Customer Lifetime Value*?**

Esta pregunta nace de lo expuesto por (Mosaddegh, 2021)

2. **¿Cuáles son los “consumidores estrella” de la compañía basándose en un análisis *Customer Lifetime Value* y categorizándolos a través de un modelo *Recency Frequency Monetary*?**

La pregunta fue inspirada en el artículo publicado de (Monalisa, 2019) donde se conversa de la importancia de definir a los consumidores estrellas para evitar invertir mucho en clientes que no ofrecen mucho retorno para un negocio y maximizar el buen servicio provisto a los consumidores más leales.

Análisis de situación laboral actual

Se pretende realizar un análisis acerca de la situación laboral actual de las 10 mejores empresas del país:

1. Análisis de rotación de personal entre empresas
2. Perfiles de empleados que buscan las 10 mejores empresas de Ecuador

Área propuesta

Sector económico – Área laboral

Métodos de Recolección

El procedimiento para obtener los datos sería a través de los datos publicados por la Superintendencia de compañías, de la cual se extraerá las 10 mejores empresas del país.

Una vez obtenido los nombres de las empresas se procederá a realizar web scraping a los perfiles de LinkedIn de las diferentes compañías, para obtener datos de sus trabajadores como profesión, nivel de educación, instituciones donde estudiaron, habilidades que poseen, cursos realizados, empresas anteriores, etc. Para poder realizar un análisis de rotación de personal entre empresas

Descripción del dataset

El conjunto de datos incluirá los siguientes campos:

1. Nombre de la persona
2. Habilidades
3. Cursos realizados
4. Empresas anteriores
5. Profesiones
6. Universidad de formación

7. Carrera
8. Nivel de formación
9. Edad
10. Empresa actual
11. Lugar de residencia
12. Tiempo que lleva trabajando en la empresa

Preguntas de negocio

- 1. ¿Cuáles son los perfiles profesionales más adecuados para ocupar cargos dentro de una de las 10 mejores empresas de Ecuador?**

Esta pregunta nace de encontrar herramientas útiles en los artículos publicados por (Funatsu,2011) (Papoutsoglou,2017) y (James,2018).

- 2. A partir de un perfil profesional ¿Cómo se ve el perfil ideal para ocupar un determinado cargo dentro de estas empresas?**

La pregunta nace del análisis de (Roca, 2019).

Repositorio de Github

<https://github.com/axvargas/Facts-and-Stats>

Bibliografía

Roca, T. (2019). *Identifying AI talents among LinkedIn members*.

Funatsu, K., & Hasegawa, K. (2011). *Knowledge-oriented applications in data mining*. In-Tech.

Monalisa, S., Nadya, P., & Novita, R. (2019). Analysis for Customer Lifetime Value

Categorization with RFM Model. *Procedia Computer Science*, 161, 834–840.

<https://doi.org/10.1016/j.procs.2019.11.190>

Mosaddegh, A., Albadvi, A., Sepehri, M. M., & Teimourpour, B. (2021). Dynamics of customer segments: A predictor of customer lifetime value. *Expert Systems with Applications*, 172, 114606. <https://doi.org/10.1016/j.eswa.2021.114606>

M. Papoutsoglou, N. Mittas and L. Angelis, "Mining People Analytics from StackOverflow Job Advertisements," 2017 43rd Euromicro Conference on Software Engineering and Advanced Applications (SEAA), 2017, pp. 108-115, doi: 10.1109/SEAA.2017.50.

Andrew James Clements & Caroline Kamau (2018) Understanding students' motivation towards proactive career behaviours through goal-setting theory and the job demands–resources model, *Studies in Higher Education*, 43:12, 2279-2293, DOI: 10.1080/03075079.2017.1326022