Sell That Twice Business Analysis

By: Juan Herrera

Abstract

The goal of this project is to analyze historical sales data as well as customer data from an e-commerce store in Australia in order to learn more about different buying patterns from customers based on demographics. I found the data in kaggle where four different csv files about the store are available. I had to perform some joins before starting with my analysis. My analysis is the previous step to what it will then become a data science solution, my analysis is helping us learn more about our customers with the hope that this will allow us to be better at customer segmentation and therefore increase customer retention

Design

Many businesses of any size are now making data-driven decisions and this is because they have enough data to make this happen. If I work in a customer-focused company these are analyses that are definitely part of the decisions being made. As previously mentioned, I took the data from kaggle where the goal is to use the data to understand what product was sold the most last month but I took the analysis a little further.

Data

After performing the join on all my tables, I ended up with 5000 rows of data which is perfect for Excel. I have great features such as age, state in Australia where the order came from, gender, total price (which is the amount of money people paid) and delivery date.

Algorithms

Feature engineering

1. Performed one-to-many joins in order to bring all the data together

2. Created new columns such as “days elapsed” by subtracting the days in which the item was delivered minus the day in which the item was ordered, that gave me the number of days elapsed between order and delivery
3. Created different pivot tables to understand and aggregate my data
4. Used tableau to visualize data through dashboards

Tools

- Google sheets and Microsoft Excel
- Tableau
- Pandas (I used pandas to perform the joins)

Communication