Superstore Sales Dataset

Predict Sales using Time Series

Abstract

THE SUPER STORE DATASET CONTAINS DATA ON ORDER DETAILS OF 1 CUSTOMERS AND STATISTICS ON THE PRODUCTS PURCHASED, THE CITIES AND THE DATE ON WHICH THEY WERE PURCHASED

Design

This project originates from the Data Science Bootcamp (T5) to find which cities sales highest by using the dataset through exploratory data analysis.

Data

Superstore sales dataset contain 18 features and 9800 unique values.

A few feature highlights include the row id ,order id, order date , ship date, ship mode, customer id, customer name, segment , country , city , state, postal code , region, product id , category, sub category, product name , sales

Algorithms

Main Feature Superstore Sales :

Ship Date : Product shipping date.

Ship Mode : pattern of shipped the product.

Segment : Type of customer.

Postl Code : Number of postcode.

Sales_level : Type of sales :costly, medium, cheap.

Preprocessing:

- Remove null value and check if there are any missing value
- Convert data type of some feature
- Add some columns for using function

Visualization:

Using the Matplotlib and Seaborn to show the Top 10 most purchased product and most purchased city.

Tools

- Numpy and Pandas for data manipulation.
- Matplotlib and Seaborn for plotting.

Communication

Presentation and Visualization.