Group Project Midterm

🎯 **The Goal:** Create an app that solves a business problem of your choice

🚀 **Team:** 2-3 students

🎙️ **Speech:** 21st of October

⭐ Tasks' Descriptions

The goal of this Group Project is to create a small app in python and streamlit that solves a business or society problem using a linear regression. You will have to structure your app as follow ➡️

- Create a page in Streamlit with the Business Case Presentation and the Data Presentation
- Create a page in Streamlit with the Data Visualization to explain relevant insights on dataset
- Create a page in Streamlit with model prediction that predict a key variable and solve a problem.

The presentation will be a 8 min presentations were all team members should speak. You will use your app as a support to your speech.

⏳ Timeline

01 - Set Teams - Find Dataset/Problem to solve - 07/10

02 - Build Landing Page describing the data - Monday 09/10

03 - Build Data Visualization Page - Wednesday 14/10

04 - Build the forecast page - Monday 16/10

06 - Speech Day - Wednesday 21/10

🛠️ Technologies

➡️ The structure of the app: **Streamlit**

➡️ The data loading, processing : **NumPy and Pandas**

➡️ The visualization part: **Looker, Seaborn, Matplotlib**

➡️ The Model Part: Linear Regression with **Scikit-Learn**

🥇 Grading Distribution

- 60 % on the App
  - Introduction page describing the problem students are trying to solve and presenting the dataset
  - Visualization page describing interesting insights in the dataset
  - Prediction page making a prediction on one of the key variables of the dataset and solving a problem
- 40 % based on in Class Presentation

💡 Inspirations

01 Data source

Find Open Datasets and Machine Learning Projects | Kaggle

Download Open Datasets on 1000s of Projects + Share Projects on One Platform. Explore Popular Topics Like Government, Sports, Medicine, Fintech, Food, More. Flexible Data Ingestion.

https://www.kaggle.com/datasets

Dataset Search

https://datasetsearch.research.google.com/

There are 25 csv datasets available on data.world.

Find open data about csv contributed by thousands of users and organizations across the world.

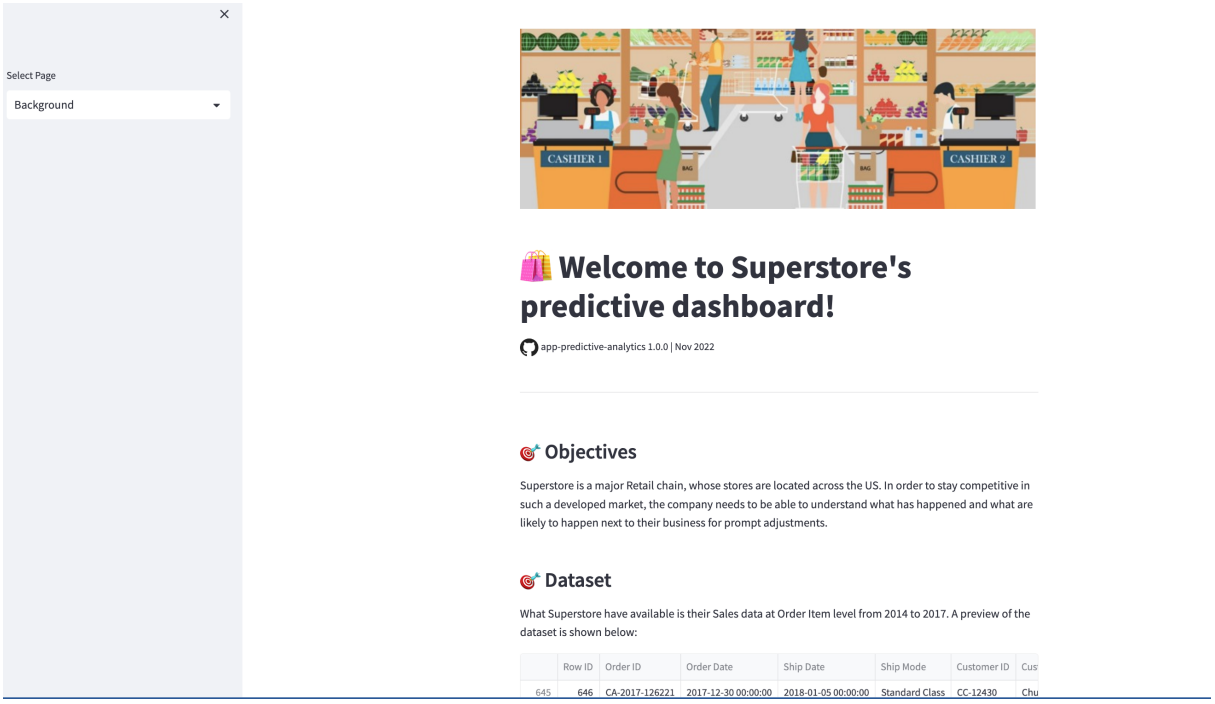
https://data.world/datasets/csv

02 APP

link1:

salesPredApp 2.zip

https://bichpham102-predanalyticsapp-2-main-1f8r4b.streamlit.app/



🎯 Objectives

Superstore is a major Retail chain, whose stores are located across the US. In order to stay competitive in such a developed market, the company needs to be able to understand what has happened and what are likely to happen next to their business for prompt adjustments.

🎯 Dataset

What Superstore have available is their Sales data at Order Item level from 2014 to 2017. A preview of the dataset is shown below:

	Row ID	Order ID	Order Date	Ship Date	Ship Mode	Customer ID	Cus
645	646	CA-2017-126221	2017-12-30 00:00:00	2018-01-05 00:00:00	Standard Class	CC-12430	Chu
906	907	CA-2017-143259	2017-12-30 00:00:00	2018-01-03 00:00:00	Standard Class	PO-18865	Patr
907	908	CA-2017-143259	2017-12-30 00:00:00	2018-01-03 00:00:00	Standard Class	PO-18865	Patr
908	909	CA-2017-143259	2017-12-30 00:00:00	2018-01-03 00:00:00	Standard Class	PO-18865	Patr
5,091	5,092	CA-2017-156720	2017-12-30 00:00:00	2018-01-03 00:00:00	Standard Class	JM-15580	Jill

Source: https://www.kaggle.com/datasets/vivek468/superstore-dataset-final

📊 Superstore's Performance Prediction

Select start date

2014/01/03

Pick a period to forcast (months)

24

0 48

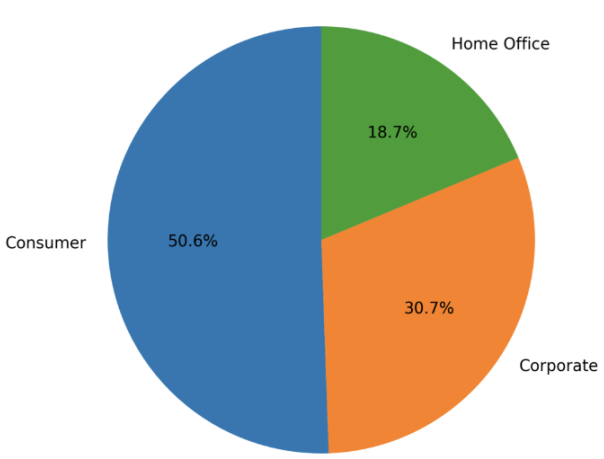
Superstore's Monthly Sales Value in $

Details

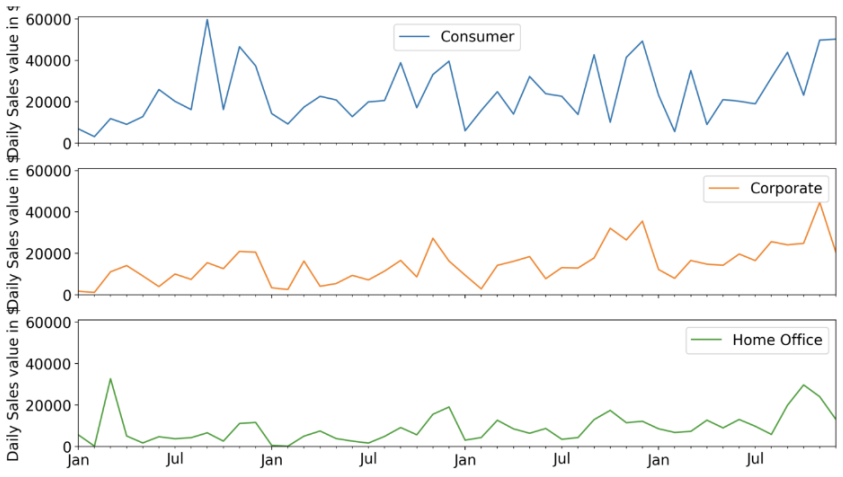
Current Performance | Forecast

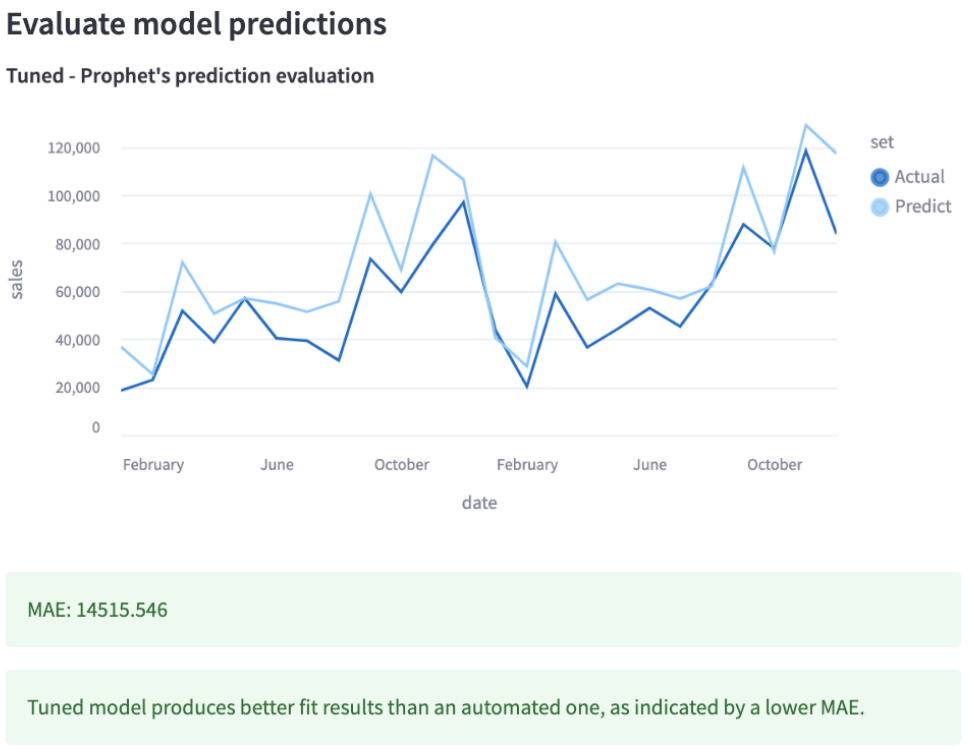
by Component | by Segment | by Category

Sales value contribution by Segment from 2014-01-03 to 2017-12-30.



Daily Sales value by Segment from 2014-01-03 to 2017-12-30.





Link2:

https://gaetanbrison-linearregressionapp-ml-project-2wc4q7.streamlit.app/

code

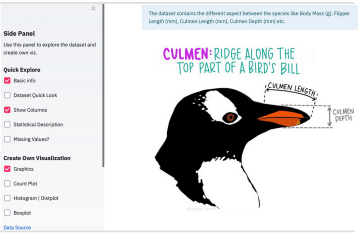
https://github.com/NYU-DS-4-Everyone/Linear-Regression-App

03 Useful Links

Build Your First Data Visualization Web App in Python Using Streamlit

Self-exploratory visualization using Streamlit visual expression's power in less than 20 minutes

https://towardsdatascience.com/build-your-first-data-visualization-web-app-in-python-using-streamlit-37e4c83a85db



Streamlit | Create Interactive Dashboards With Streamlit

Streamlit is the fastest growing MI and data science dashboard building platform. Learn how to create interactive dashboards with streamlit

https://www.analyticsvidhya.com/blog/2020/10/create-interactive-dashboards-with-streamlit-and-python/

Congrats you made it 🎉