Programming for AI - TABA

Project Journal

SANDEEP KUMAR (x23282835), x23282835@student.ncirl.ie

Furniture Sales Forecasting Using Time Series

Project Title: Exploring Time Series Algorithms to Forecast Sales Targets in Retail Sector

Objective: To develop a furniture sales forecasting model using time series analysis, using ARIMA and Prophet models for accurate and actionable predictions.

Streamlit Dashboard: In the last section, I have added screenshots of my streamlit app dashboard for this project which have two tabs. Home and EDA. Home section is having monthly forecasted graph and also having tables showing monthly forecasted values which have confidence intervals also. Monthly forecasted value is dynamic which can be modified and forecast can be generated on the page. I have created this using trained prophet model. EDA tab is having all the missing values, duplicate rows, word cloud, histogram all kinds of graph which I have created using Data Profiling library.

Furniture Sales Forecasting Dashboard: https://ts-dashboard-ft317gdnpk.streamlit.app/

Please find the project journal for furniture sales forecasting in tabular format and also in details.

Project Journal in Table Format

Date	Phase	Activities	Challenges	Resolution
28-11-2024	Problem Understanding and Dataset search	Brainstorming ideas for the project. Searching for common theme and suitable dataset.	Finding dataset on common theme was very difficult due to limited access to clean and structured retail sales data.	After brainstorming got the idea of Retail Sales forecasting. I have chosen dataset which is publicly available. Dataset is furniture sales forecasting
05-12-2024	Data Collection and Data Loading	Loading data using pandas libraries	There was issue while loading csv file as it was giving encoding errors.	After searching the error in internet found out that encoding should be added as parameters and values should be “latin1”. encoding = “latin1”

07-12-2024	Data Pre-processing	Checked dataset distribution, outliers, missing values, duplicate rows	Dataset was left-skewed, outliers were present	Used log-transformation for skewness, and IQR method to remove outliers from the dataset
09-12-2024	Data Preparation For Time Series	Analyse columns to find out manually which have date time	Two columns – ["Order Date" And "Ship Date"] were there whose datatype was "object"	I converted date columns to datetime format. After converting data type of these columns became "datetime64[ns]"
12-12-2024	Exploratory Data Analysis (EDA)	Performed time series decomposition, generated ACF and PACF plots to identify lags, and tested stationarity using the Augmented Dickey-Fuller test.	Noise and seasonality made interpreting trends challenging.	Applied smoothing techniques and seasonal decomposition to isolate key patterns.
15-12-2024	Time Series Analysis	Stationarity check using Augmented Dickey-Fuller test. , ACF & PACF plot identify lags, Time series Decomposition plots	Data was not resampled to monthly frequency, Noise and seasonality were present	Resampled data to monthly frequency, and visualized trends and seasonality. Applied smoothing techniques and seasonal
17-12-2024	Model Building	Trained ARIMA [Auto-ARIMA & SARIMAX] and Prophet models; Optimized parameters using Auto-ARIMA; incorporated seasonal and trend adjustments in Prophet	I got issues while running prophet model. Also, ARIMA required significant adjustments to achieve stationarity, and Prophet needed fine-tuning for seasonality.	I found out that library name for prophet has changed from fprophet to prophet. I Used differencing and log transformations for ARIMA and then adjusted seasonality parameters for Prophet.
20-12-2024	Model Evaluation and Comparison	Evaluated models using MAE, MSE, MAPE, and R^2, visually and quantitatively compared forecasts and also created a structured benchmarking table.	It was difficult to compare results from ARIMA and Prophet due to differing assumptions.	Prophet got good results in terms of metrics because I used seasonality mode which takes care of seasonality and hence got better result with prophet

21-12-2024	UI Creation	Create Dashboard showing forecasted graph and table for 12 months and other tab for EDA using Data Profiling	I got issues in installing libraries for "pandas_profiling", Data Profiling was taking so much time to load at every refresh.	I found that libraries name has changes from "pandas_profiling", to "ydata_profiling". I then added cache to avoid reloading of EDA report.
22-12-2024	Flask API	Load trained model to be used in Flask API, test API locally	I got file not found error for trained model, date time format issue as it was adding time stamp automatically when running locally	I found that path of trained model which was in pickle format was wrong and then provided correct path, I formatted date to "forecast['ds'].dt.date" to remove unnecessary timestamp.
23-12-2024	Model Integration	Integrate Flask API with Streamlit APP	I was getting JSON error while running Streamlit App after running Flask API	I first tried testing Flask API by giving right JSON format input and was getting correct forecasted output then I found that I missed to provide Flask API URL in streamlit APP after adding it was working fine
26-12-2024	Model Deployment on Cloud	Deploy Flak API on public cloud [render] and streamlit app on streamlit cloud	While deploying Flask API, got libraries version issues for python, gunicorn and other libraries, flask was not accepting request from streamlit app, model path was not found, while deploying streamlit on streamlit cloud again was getting JSON error	I removed version from requirements.txt file and then render web service automatically downloaded compatible version. I have to add flask CORS then flask was able to accept request from streamlit app, I added "from pathlib import Path" then it was able to found my model in pickle format. I found that I didn't remove local path of Flask API URL, after adding new Flask API path which was running on cloud, it was running smoothly.

Detail Summary:

Date: 28-11-2024

Phase 1: Problem Understanding and Dataset search:

- Brainstorming ideas for the project.
- Searching for common theme and suitable dataset.
- I have to defined the forecasting horizon as 12 months.
- After brainstorming got the idea of Retail Sales forecasting.
- I have chosen dataset which is publicly available. Dataset is furniture sales forecasting

Outcome:

- After brainstorming got the idea of Retail Sales forecasting.
- I have chosen dataset which is publicly available. Dataset is furniture sales forecasting

`

Date: 05-12-2024

Phase 2: Data Pre-processing
Activities:

- Addressed missing values in the dataset. Data Collection and Data Loading
- Checked dataset distribution, outliers, missing values, duplicate rows
- Analyse columns to find out manually which have date time

Outcome:

- I got cleaned dataset which was ready for time series modeling.
- I found out that encoding should be added as parameters and values should be "latin1". encoding = "latin1" while loading csv file.
- I used log-transformation for skewness, and IQR method to remove outliers from the dataset
- I converted date columns to datetime format. After converting data type of these columns became "datetime64[ns]"

Date: 12-12-2024

Phase 3: Exploratory Data Analysis (EDA)
Activities:

- I performed time series decomposition, generated ACF and PACF plots to identify lags.
- I generated ACF and PACF plots to look significant lags.
- I conducted stationarity tests using the Augmented Dickey-Fuller method.

Outcome:

- I got clear insights into the dataset's trends, seasonality, and stationarity characteristics.
- Applied smoothing techniques and seasonal decomposition to isolate key patterns.
-

Date: 17-12-2024

Phase 4: Model Building
Activities:
ARIMA (SARIMAX):

- I identified optimal parameters using Auto-ARIMA.
- Then trained the model on historical sales data.
- And finally generated forecasts for the next 12 months.

Prophet:

- I trained the Prophet model, which take care of seasonal and trend components.
- I then produced sales forecasts for the next 12 months.

Outcome:

- I trained two models (ARIMA and Prophet) which is ready for evaluation.

Date: 20-12-2024

Phase 5: Model Evaluation and Comparison
Activities:

- I then evaluated both models using performance metrics: MAE, MSE, MAPE, and R^2.
- I also visualized and compared forecasts from ARIMA and Prophet.
- Please find my benchmarking table for quantitative comparison.

```
Model Benchmarking Results:
     Model  R-squared          MAE           MSE       MAPE
0    ARIMA   0.109030  5872.399678  7.230446e+07  43.735713
1  Prophet   0.932059  1877.451707  5.513546e+06  15.859204
```

Outcome:

- Prophet model outperformed SARIMAX with R^2 value of 0.93 and MAPE value of 15.85% which indicates better accuracy and fit to this data. MSE and MAE also showed reasonable accuracy.

Date: 26-12-2024

Phase 6: Model Deployment
Activities:

- I saved trained Prophet model in pickle format.
- Developed scripts to load models and generate forecasts.

Outcome:

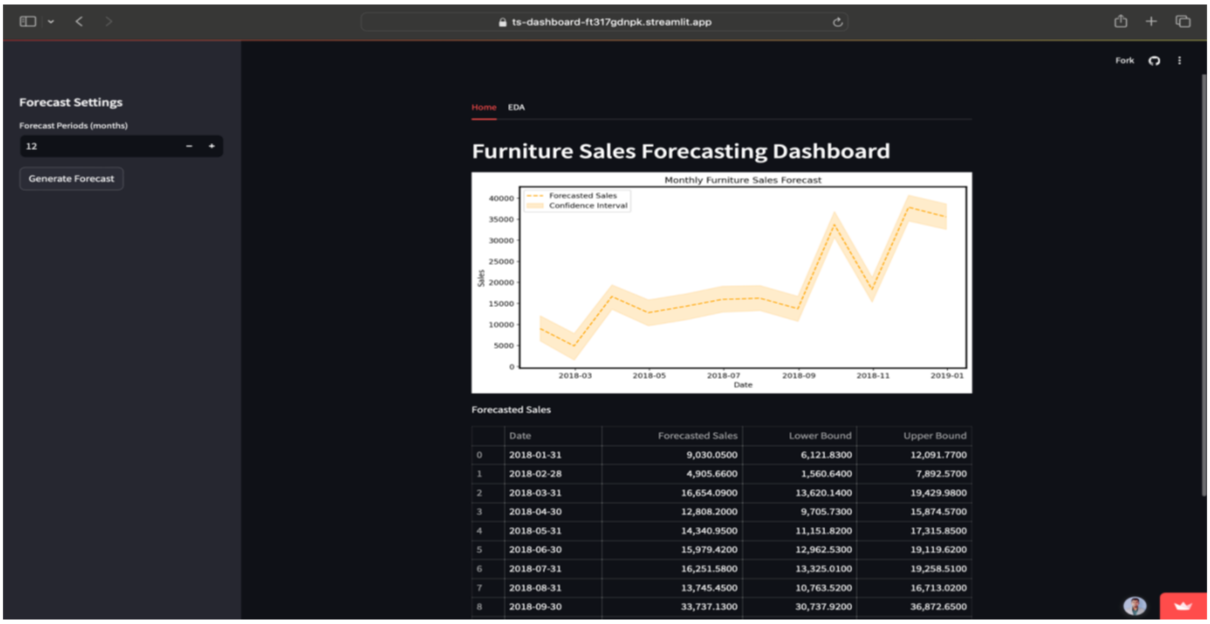
- First I deployed Flask API to public cloud (Render Cloud)
- I then deployed forecasting models to streamlit cloud.

Dashboard Details(URL) of my deployed model to Streamlit Cloud:

Furniture Sales Forecasting Dashboard: https://ts-dashboard-ft317gdnpk.streamlit.app/

Please find below screenshot of my deployed model on streamlit cloud:

STEP 1: Please visit above dashboard URL, you will be able to find below homepage.



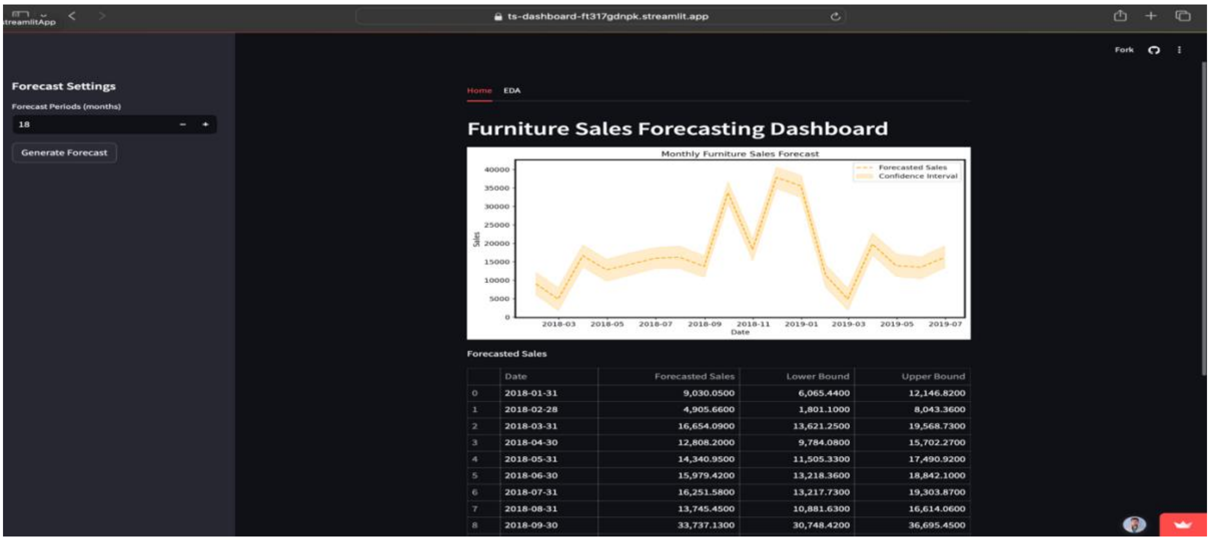
In homepage, you can see Forecast period which you can adjust and then when you will click on generate forecast, it will give you graph showing monthly furniture sales of that period which you have selected together with tabular format forecast in values with confidence interval as lower and upper bound of yhat.

Below figure shows tabular format forecast:

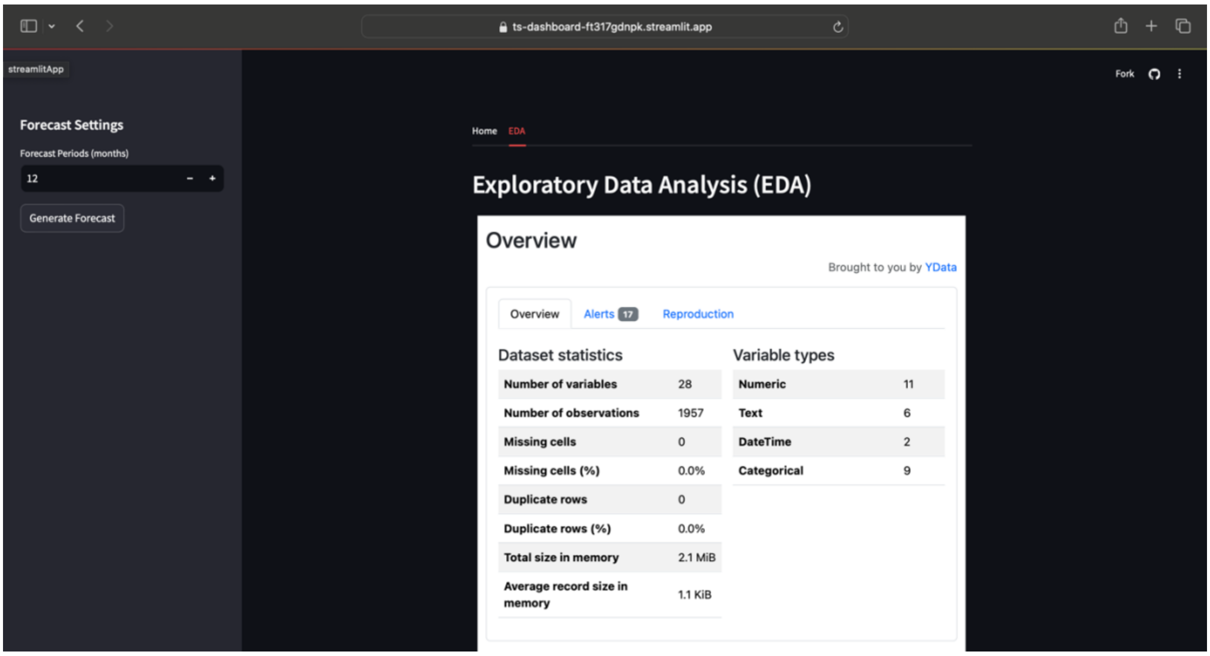
Forecasted Sales

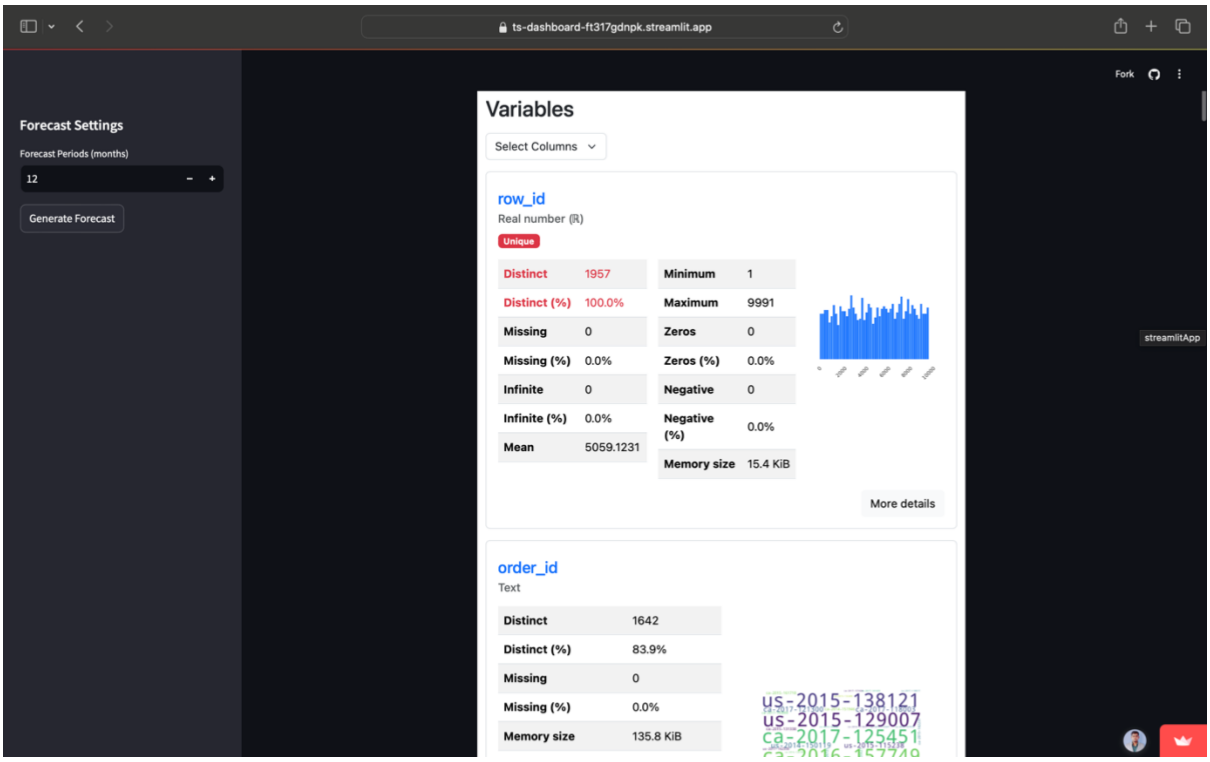
	Date	Forecasted Sales	Lower Bound	Upper Bound
0	2018-01-31	9,030.0500	6,121.8300	12,091.7700
1	2018-02-28	4,905.6600	1,560.6400	7,892.5700
2	2018-03-31	16,654.0900	13,620.1400	19,429.9800
3	2018-04-30	12,808.2000	9,705.7300	15,874.5700
4	2018-05-31	14,340.9500	11,151.8200	17,315.8500
5	2018-06-30	15,979.4200	12,962.5300	19,119.6200
6	2018-07-31	16,251.5800	13,325.0100	19,258.5100
7	2018-08-31	13,745.4500	10,763.5200	16,713.0200
8	2018-09-30	33,737.1300	30,737.9200	36,872.6500
9	2018-10-31	18,344.3300	15,331.0500	21,201.3700
10	2018-11-30	37,863.2900	34,602.6700	40,738.6400
11	2018-12-31	35,570.2400	32,613.7100	38,667.2500

STEP 2: Below figure shows snapshot when I forecasted for 18 months:

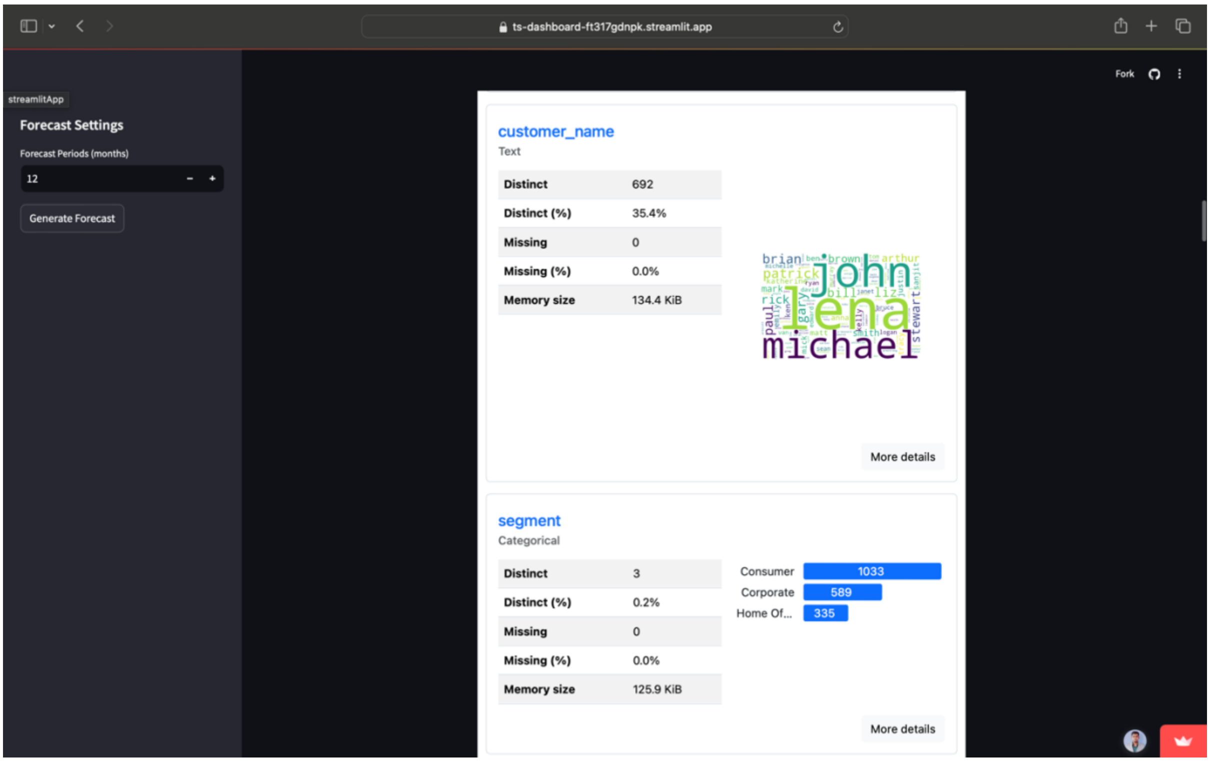


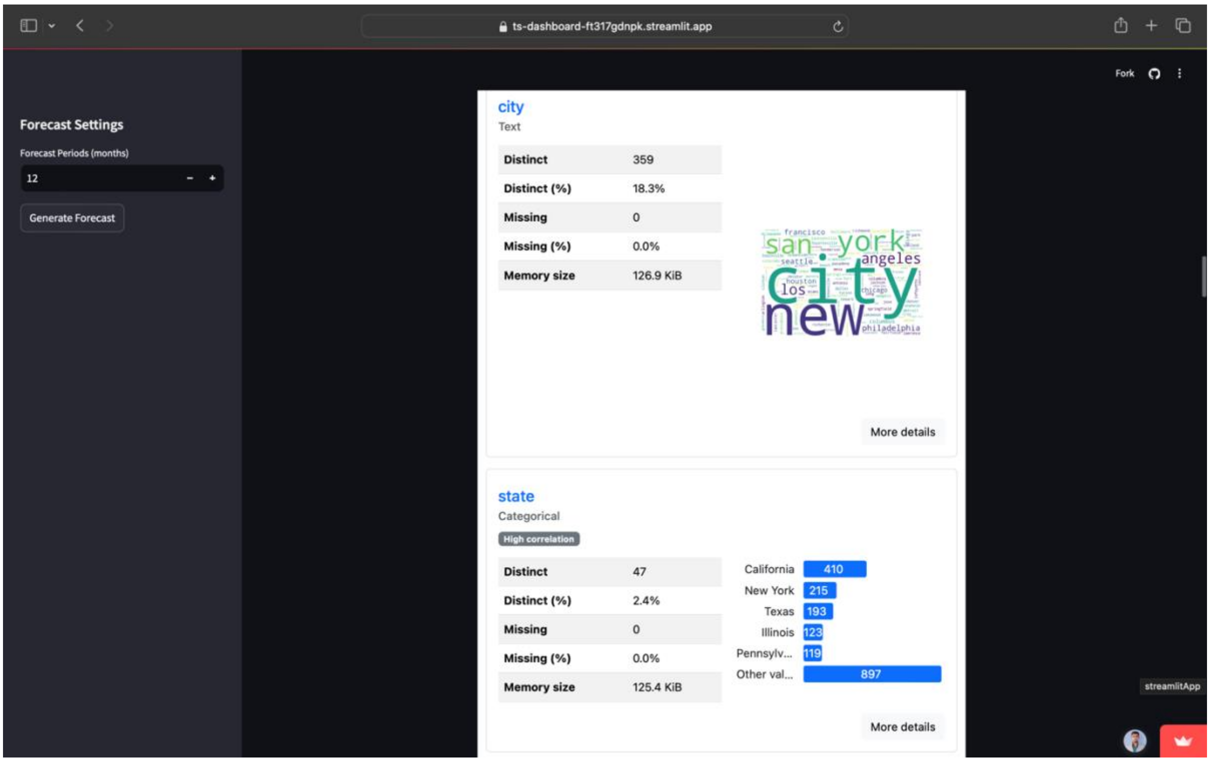
STEP 3: Below image shows EDA tab which have all types of data statistics, visualization graph, word cloud and almost all types of plot.





Below images are showing word cloud of customer_name & City.




ts-dashboard-ft317gdnpk.streamlit.app
Forecast Settings
Forecast Periods (months)
12
Generate Forecast
Fork
city
Text
Distinct 359
Distinct (%) 18.3%
Missing 0
Missing (%) 0.0%
Memory size 126.9 KiB
More details
state
Categorical
High correlation
Distinct 47
Distinct (%) 2.4%
Missing 0
Missing (%) 0.0%
Memory size 125.4 KiB
California 410
New York 215
Texas 193
Illinois 123
Pennsylv... 119
Other val... 897
More details
streamlitApp