

Investment Decision Recommendation System Documentation

Introduction:

Welcome to the Investment Decision Recommendation System documentation! This system is designed to assist users in making informed investment decisions based on various factors. Explore the detailed overview of the system's functionalities below. Built an interactive UI using streamlit and also made analysis using PowerBI.

Tasks performed :

1.Data Preprocessing :

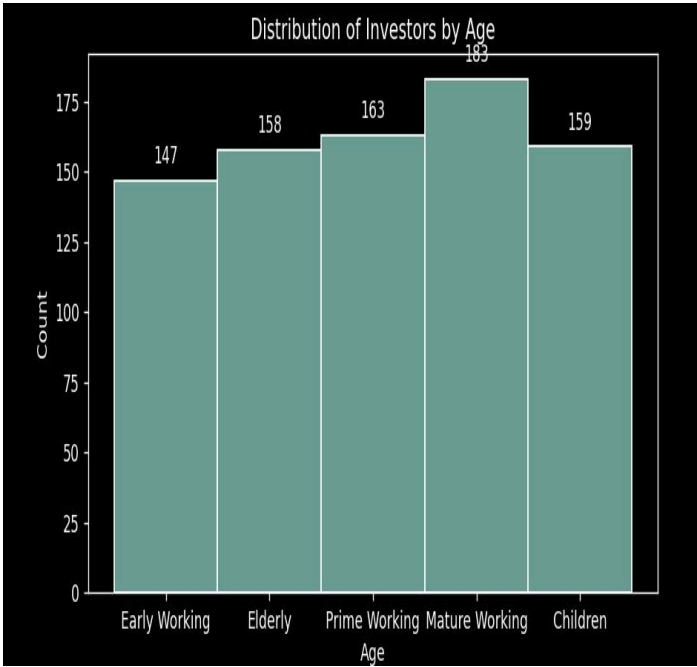
- Data Loading
The system loads the dataset containing information about individuals and their investment behavior.
- Handling Missing Values
Missing values are identified and handled using techniques like imputation or removal based on the context.
- Data Cleaning
Duplicates are removed, data types are corrected, and consistency is ensured in categorical variables.
- Encoding Categorical Variables
Categorical variables are encoded using techniques like one-hot encoding or label encoding.

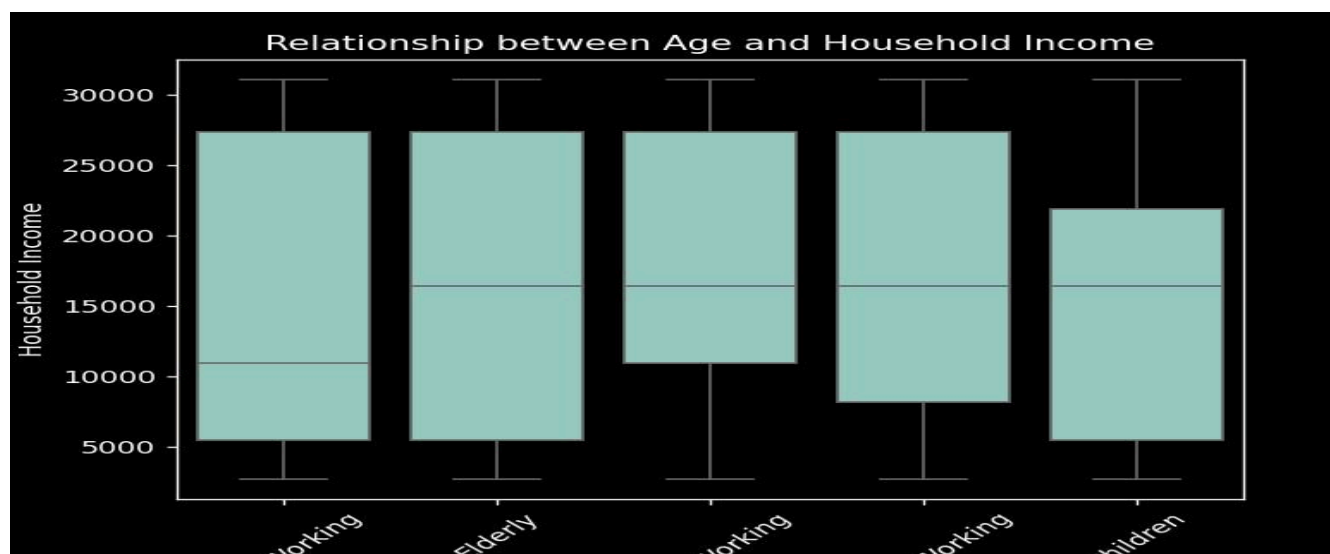
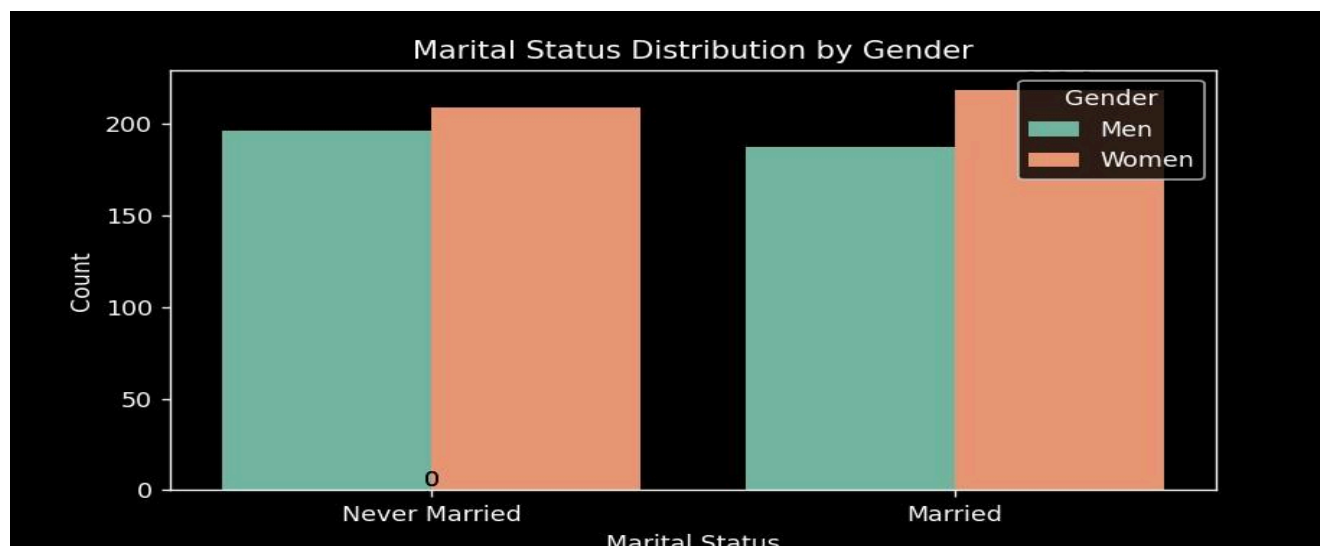
2. Visualization :

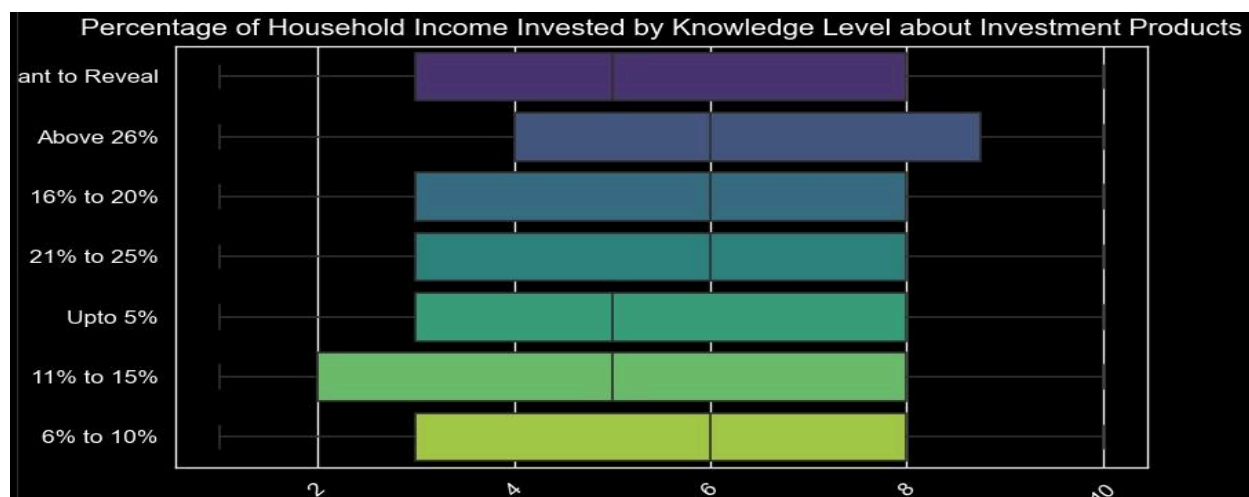
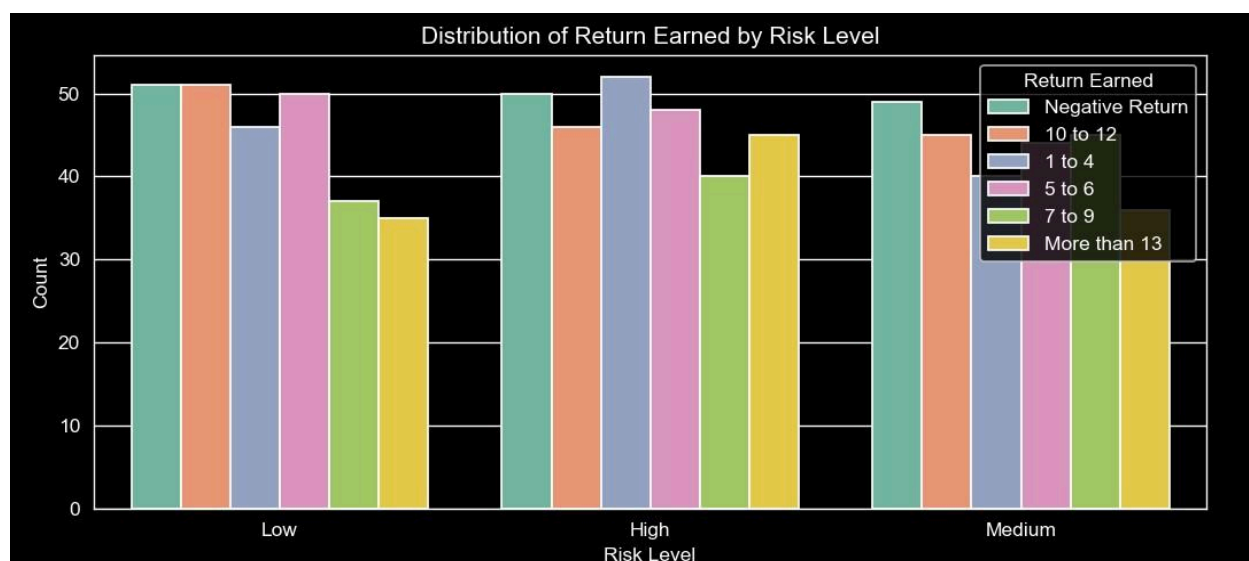
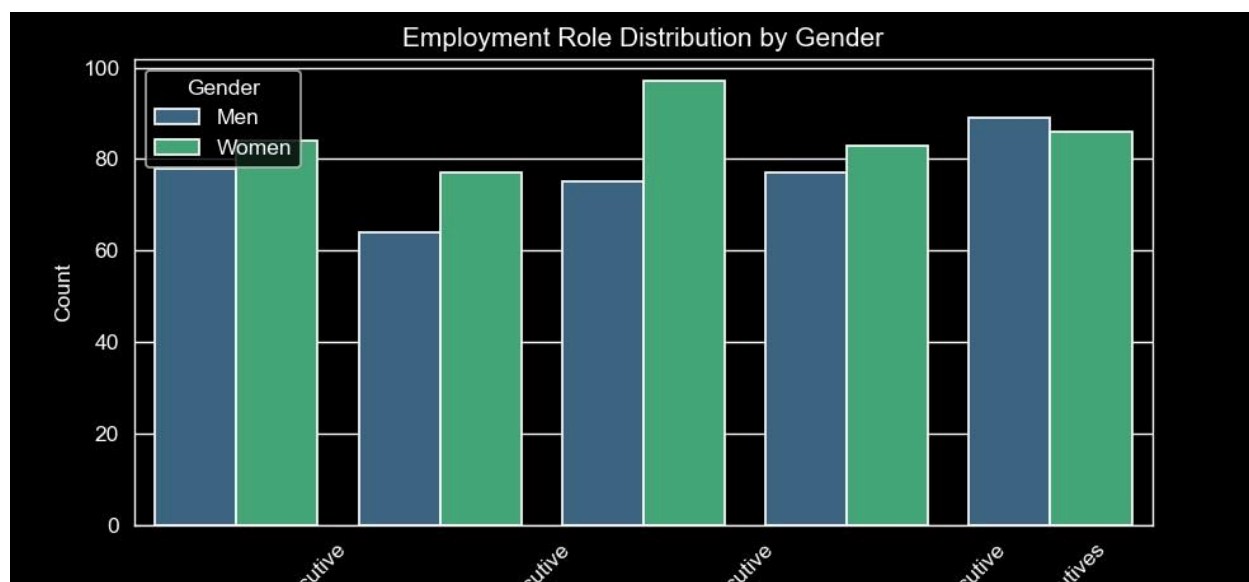
- Data Exploration
Visualizations are used to analyze demographic distribution, employment details, investment behavior insights, and investment outcomes.

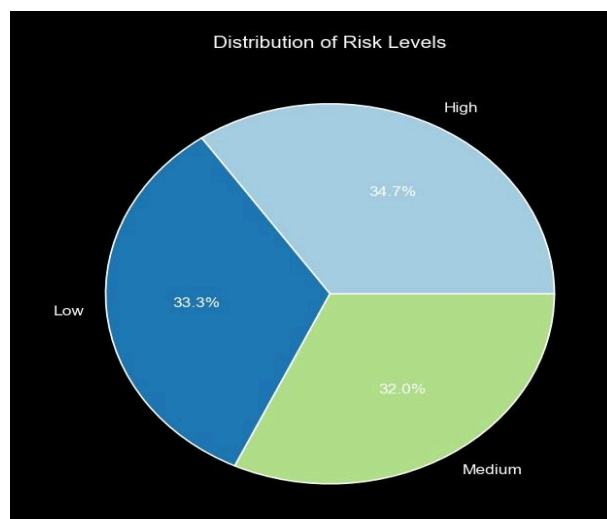
Performed many visualizations and found all the relationship among the features

PICTORIAL DEMONSTRATION





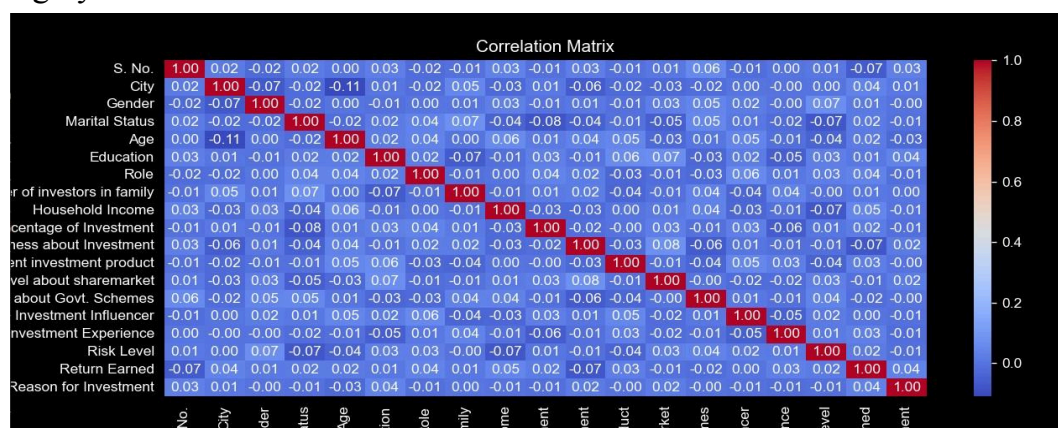




3. Best Investment Decision Identification:

- **Analyzing Correlation:**

Performed Correlation among the features to find the feature which is highly correlated



- **Baseline Model:**

Initial models such as Decision Trees , Random Forests, Neural Networks are trained to establish a performance benchmark.

- **Model Evaluation:**

The baseline model's performance is evaluated using metrics like accuracy, precision, recall, and F1-score.



Hence, the accuracy is less and underfitting we are performing models to improve the accuracy

- **Hyperparameter Tuning**

Grid search or randomized search techniques are used to tune hyperparameters and optimize model performance.

Model Comparison:

Different models, including Gradient Boosting, Neural Networks, and ensemble methods, are compared based on evaluation metrics.

- **Cross-Validation**

K-fold cross-validation is employed to validate model performance across multiple data subsets and improve accuracy

- **Final Model Selection**

The model with the highest performance and generalizability is selected as the recommendation system model.

Model Comparison

Models are compared based on evaluation metrics to identify the best-performing model for investment decision recommendation.

Here , Random Forest gave higher Accuracy and hence that is taken as the best model for investment decision recommendation.

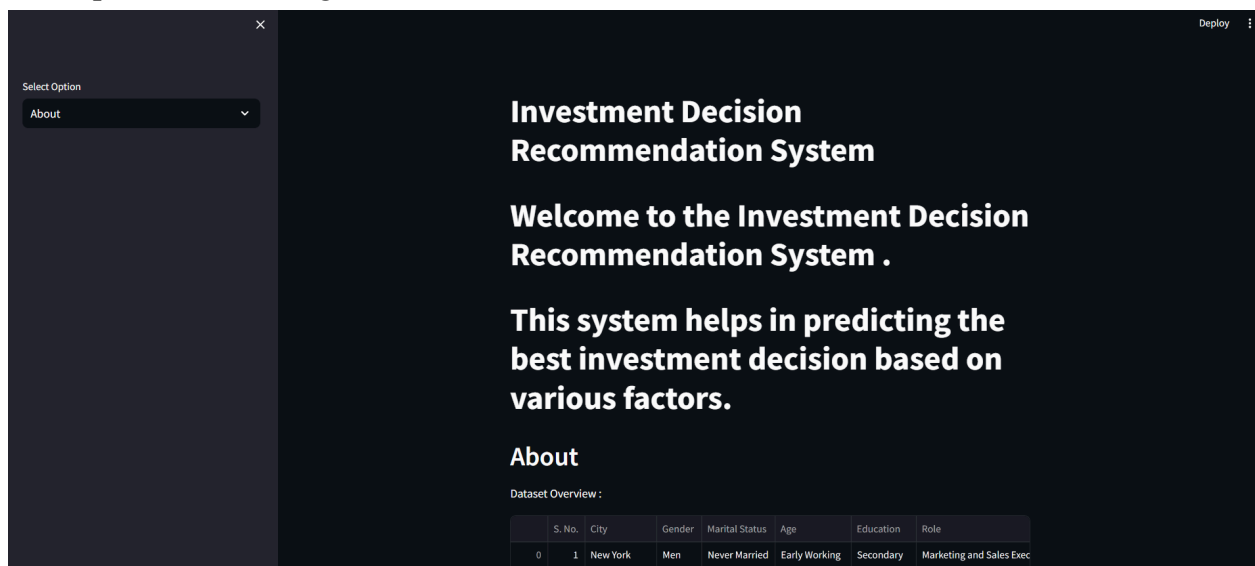
4. User Interface:

Streamlit is utilized to create a user-friendly interface for exploring data, visualizations, model inferences, and predictions.

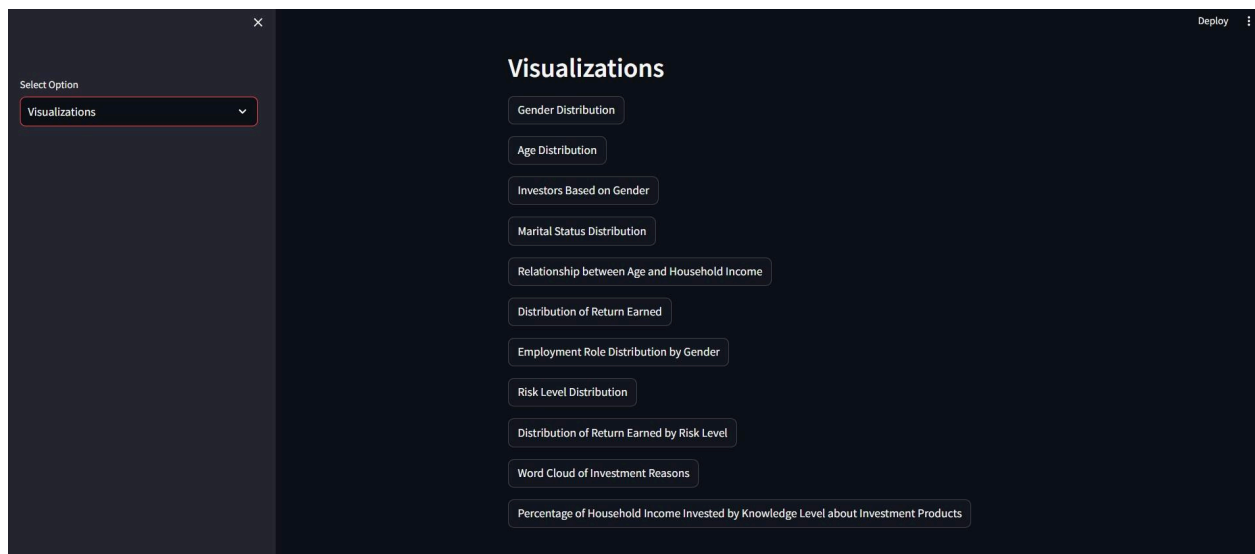
Here I have created a Interface that has a about page, Visualization page, Inference page, Prediction by user Input and a analyzing page for interactive visualization

Home Page:

A Simple Welcome Page



Visualization Page:



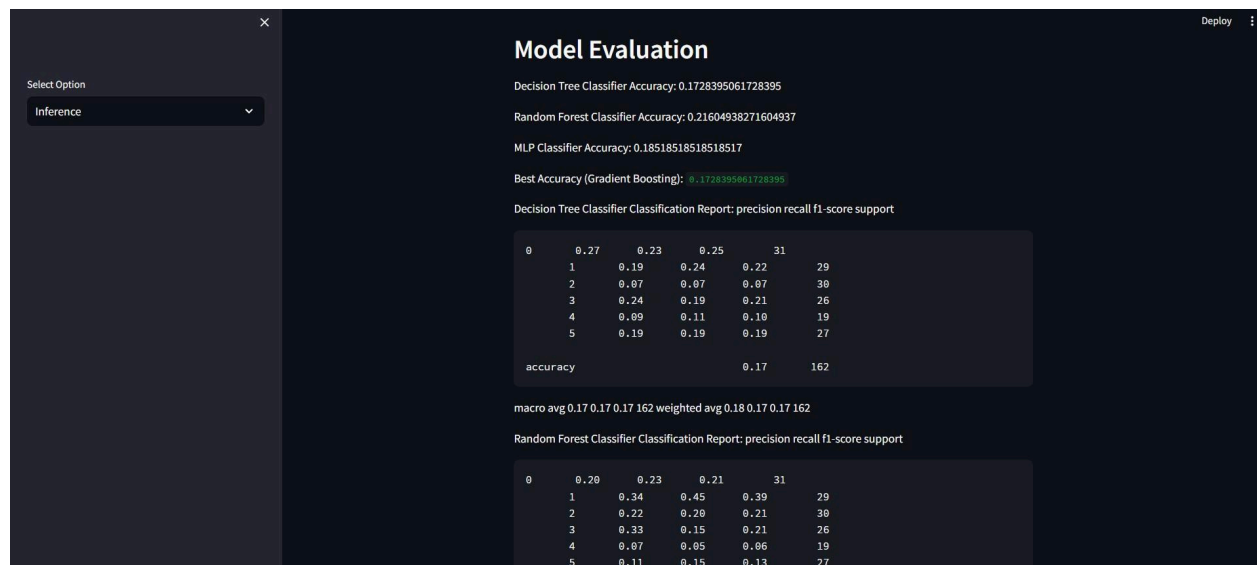
Made it user friendly and user can view the plots by click on the particular topic



An example of the selection

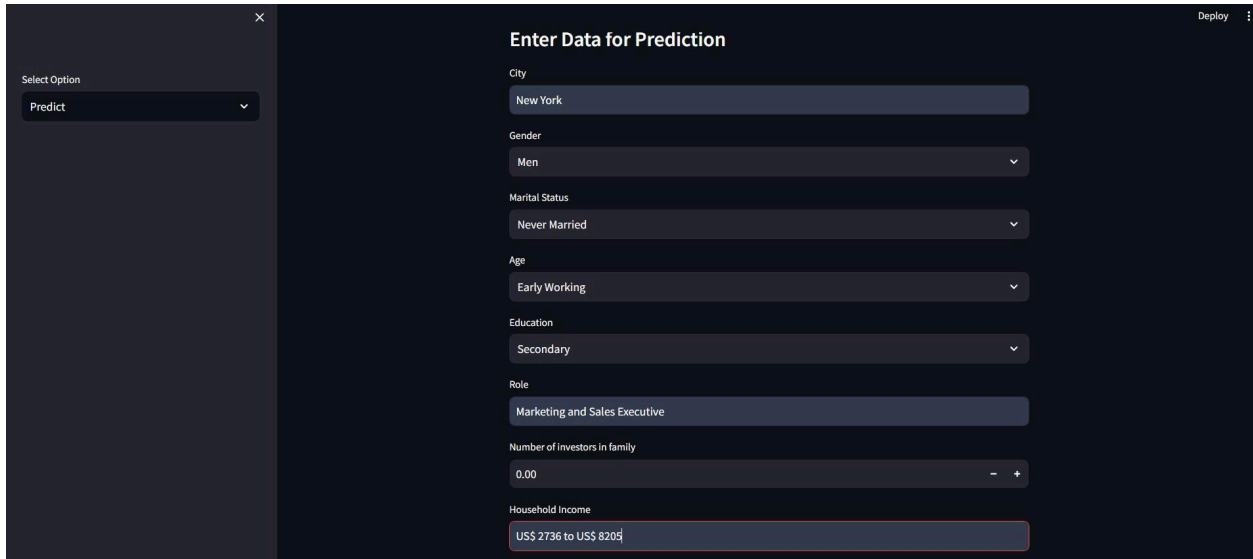
Inference Page:

In this page, we have visualized correlation and displayed the Accuracy and the classification reports



Prediction Page:

This is a very important page where users can predict the best investment decision. User have to give the input , Example is given below



The screenshot shows a web application interface for entering data for a prediction. On the left, there is a sidebar with a 'Select Option' dropdown menu currently set to 'Predict'. The main area is titled 'Enter Data for Prediction' and contains several input fields: 'City' (text input with 'New York'), 'Gender' (dropdown with 'Men'), 'Marital Status' (dropdown with 'Never Married'), 'Age' (dropdown with 'Early Working'), 'Education' (dropdown with 'Secondary'), 'Role' (text input with 'Marketing and Sales Executive'), 'Number of investors in family' (range input with '0.00'), and 'Household Income' (text input with 'US\$ 2736 to US\$ 8204'). A 'Deploy' button is visible in the top right corner.

This is the Example for the user input

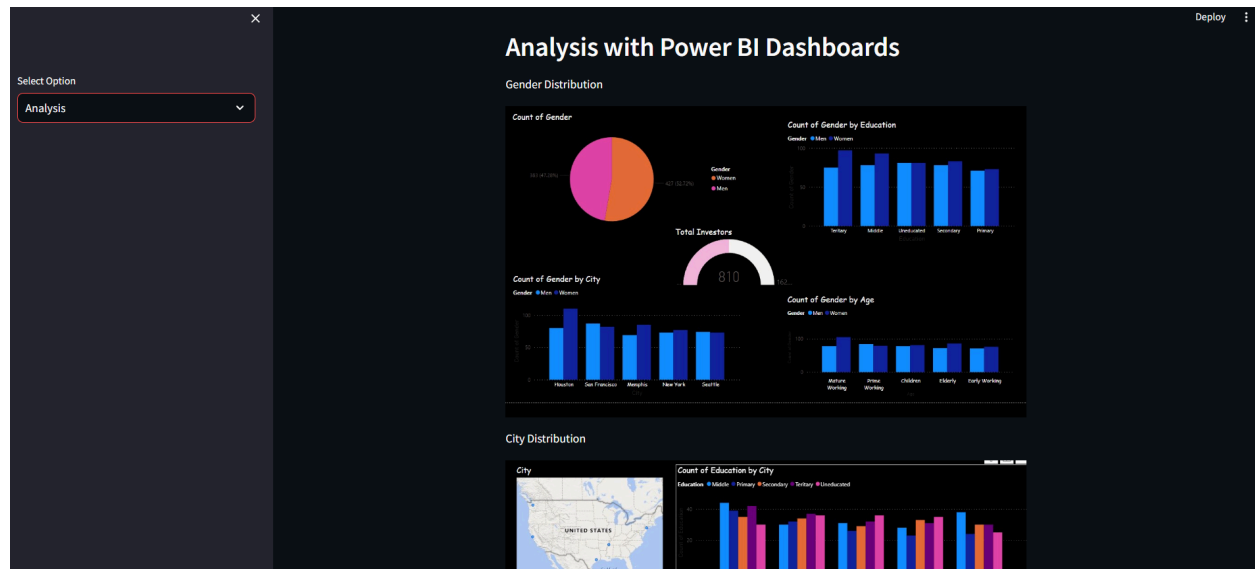
Output will be as the following:



The screenshot shows the output of the prediction. It features a series of dropdown menus for input: a top dropdown with '6', 'Select Investment Influencer' with 'Family Reference', 'Select Investment Experience' with 'Less Than 1 Year', 'Select Risk Level' with 'Low', and 'Select Reason for Investment' with 'Tax'. Below these, a horizontal line separates the inputs from the output, which is 'Predicted returns=7 to 9'. A 'Deploy' button is in the top right corner.

Analysis Page:

This page is created to analyse the interactive visualization done by the PowerBI



5. Furthermore for interactive insights created a PowerBI dashboard:

- Dashboard Integration

Power BI dashboards are integrated into the system to provide interactive and insightful visualizations.

Power BI dashboards analyze gender distribution, city distribution, return earned distribution, knowledge about investment, and household income.

Conclusion

The Investment Decision Recommendation System leverages data preprocessing, model selection, evaluation, visualization, user interface, and Power BI analysis to assist users in making better investment decisions. The systematic approach ensures reliability, accuracy, and user-friendliness throughout the system

This document and code was purely done by me - Varsha S pursuing Data Science from PSG College Of Technology for the test conducted by Buckman Laboratories for the Internship Opportunity.

Thank You
