



AIML INTERNSHIP

HENOTIC TECHNOLOGY PRIVATE LIMITED

7th July 2022 to 06th September 2022

Presented by:

K KRISHNA CHAITANYA

$\frac{3}{4}$ CSE

Regd.No: 20B91A05E3

**DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING
SAGI RAMA KRISHNAM RAJU ENGINEERING COLLEGE**

CONTENTS

- INTRODUCTION ABOUT DATASET
- OBJECTIVE & PROBLEM STATEMENT
- DATA SCIENCE PROJECT LIFE CYCLE
- RESULTS
- CONCLUSIONS AND FUTURE WORK
- REFERENCES





DATASET:

BLACK FRIDAY SALES PREDICTION



INTRODUCTION ABOUT DATASET

- Black Friday is an informal name for the Friday following Thanksgiving Day in the United States, which is celebrated on the fourth Thursday of November.
- The major challenge for a Retail store or eCommerce business is to choose product price such that they get maximum profit at the end of the sales.
- Our project deals with determining the product prices based on the historical retail store sales data. After generating the predictions, our model will help the retail store to decide the price of the products to earn more profits.
- This dataset comprises of sales transactions captured at a retail store. It's a classic dataset to explore and expand your feature engineering skills and day to day understanding from multiple shopping experiences.

INTRODUCTION ABOUT DATASET

- The data consists of 12 columns and 537577 records.
- The data contained features like:
 - **User_ID**: Unique ID of the user. There are a total of 5891 users in the dataset.
 - **Product_ID**: Unique ID of the product. There are a total of 3623 products in the dataset.
 - **Gender**: indicates the gender of the person making the transaction.
 - **Occupation**: shows the occupation of the user, already labelled with numbers 0 to 20.
 - **City_Category**: Cities are categorized into 3 different categories 'A', 'B' and 'C'.
 - **Stay_In_Current_City_Years**: Indicates how long the users has lived in this city.
 - **Marital_Status**: is 0 if the user is not married and 1 otherwise
 - **Product_Category_1 to _3**: Category of the product. All 3 are already labelled with numbers.
 - **Purchase**: Purchase amount.(Target variable)

OBJECTIVE & PROBLEM STATEMENT

- The main agenda of this project is to Perform extensive Exploratory Data Analysis(EDA) on the Black Friday Sales Dataset.
- Build an appropriate Machine Learning Model that will help a Retail store or eCommerce business is to choose product price such that they get maximum profit at the end of the sales.

Problem Statement :

- Design a predictive model with the use of machine learning algorithms to Predict purchase amount.

DATA SCIENCE PROJECT LIFE CYCLE

1. Data Exploratory Analysis

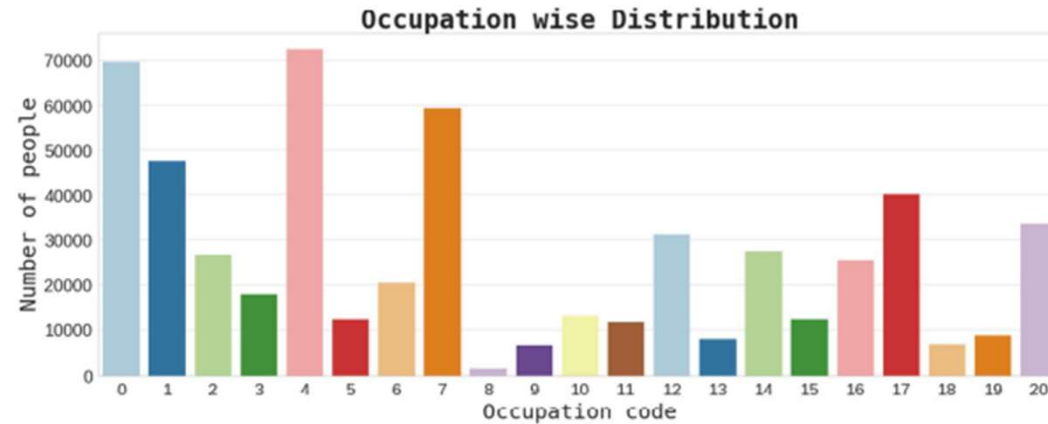
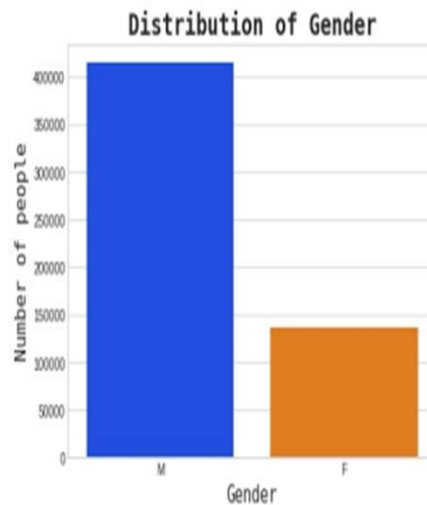
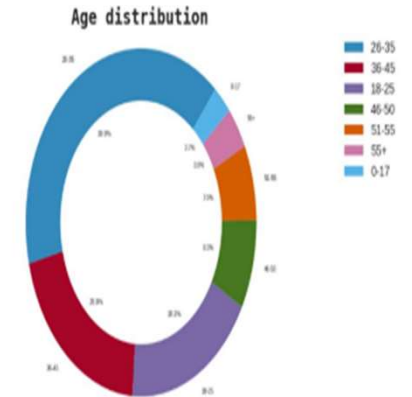


Fig A.2.2 Univariate Analysis



DATA SCIENCE PROJECT LIFE CYCLE

2. Data Pre-processing

- i. Check the Duplicate and low variation data
- ii. Identify and address the missing variables
- iii. Handling of Outliers
- iv. Categorical data and Encoding Techniques
- v. Feature Scaling

3. Selection of Dependent and Independent variable

4. Data Sampling Methods

4. Training Models



RESULTS

Based on the analysis , these are the results generated from different models

	Model Name	Mean_Absolute_Error_MAE	Adj_R_Square	RMSE	MAPE	MSE	RMSLE	R2_score
0	LinearRegression()	3596.239665	0.119808	4650.528605	68.646215	21627416.31	8.444736	0.120336
1	DecisionTreeRegressor()	2767.148383	0.372955	3925.209998	39.70573	15407273.52	8.275175	0.373331
2	(DecisionTreeRegressor(max_features=1.0, random_state=13363782))	2326.102366	0.594575	3156.231891	34.958592	9961799.752	8.057134	0.594818
3	SVR()	3792.659928	-0.014722	4993.294507	67.469226	24932990.03	8.515851	-0.014113
4	XGBRegressor(base_score=0.5, booster='gbtree', n_estimators=1000)	2215.496592	0.642746	2962.801144	34.52999	8778190.617	7.99389	0.64296
5	KNeighborsRegressor()	3134.156453	0.259693	4265.004884	57.717826	18190266.66	8.358199	0.260137
6	(ExtraTreeRegressor(random_state=13363782))	2438.97422	0.540968	3358.423464	36.126774	11279008.16	8.119227	0.541243
7	(DecisionTreeRegressor(criterion='friedman_mse', random_state=13363782))	2257.902371	0.639532	2976.096025	36.048128	8857147.547	7.998368	0.639748
8	XGBRegressor(base_score=0.5, booster='gbtree', n_estimators=1000)	2270.840622	0.635892	2991.085437	33.881512	8946592.089	8.003392	0.63611

CONCLUSIONS AND FUTURE WORK

- The model results in the following order by considering the model Adjusted_r2 ,RMSE ,MAE and MAPE
 - 1) XGB Regressor
 - 2) Extra Trees Regressor
 - 3) Decision tree Regressor
- Recommended model – XGB Regressor with Random Sampling technique as a best fit for the given BF Purchase dataset.
- XGB Regressor Model got an Adj_r_square of about 64.2% and got a RMSE value of 2692.5
- As we got the best in every parameter we consider this is the best model to find Purchase sales prediction.
- Future Work – We have a large enough dataset, so we can use neural networks such as an artificial neural network to build a model which can result in better performance.

REFERENCES

- Dataset Link : <https://www.kaggle.com/datasets/sdolezel/black-friday?select=train.csv>
- Website : <https://blackfriday.com/news/black-friday-predictions>



THANK YOU

