Advertisement Success Prediction

By Rounak Sharma

Overview

- When one starts a business or is running a business, one needs to make people aware of the fact. Promotion is important for any type of business to let others know about the business.

- Advertising is important for every aspect of a business. It plays an imperative role for both manufacturers and consumers.

- Advertising is important for the business on the whole as it lets the business gain more customers, thereby increasing business turnaround.

Problem Statement

- Building a machine learning model to predict whether an advertisement buy will lead to a net gain to the business.

- This model will predict how much an advertisement add value to the business.

Stakeholder

- **Head of The Marketing Department**
- Head of Advertising Firm
- Manager of Regulatory Bodies

Data Science Metric

- **F1 Score**

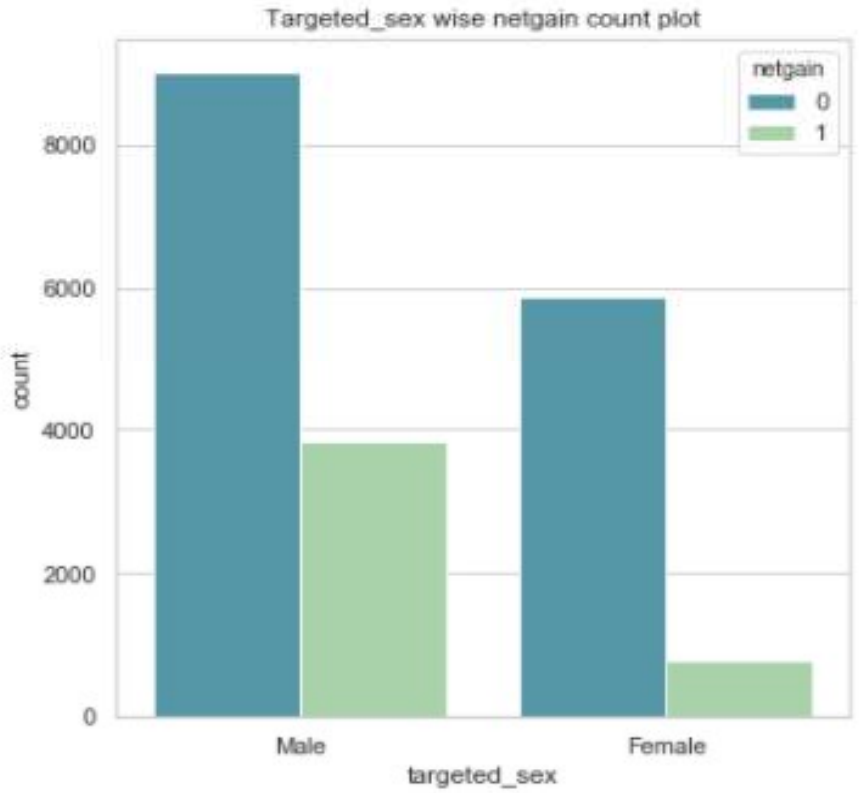
Data

Feature	Description
UserID	Unique id for each row
ratings	Ratings for advertisement
airlocation	Country of origin
airtime	Time when the advertisement was aired
average_runtime(minutes_per_week)	Minutes per week the advertisement was aired
targeted_sex	Gender targeted by advertisement
genre	Type of advertisement
industry	The industry to which the product belonged
relationship_status	Relationship status of the customers
expensive	Advertisement expenses
money_back_guarantee	Whether the product offers a refund or not
netgain	Targeted feature which identify profit or loss for ads

EDA

Targeted Sex Count

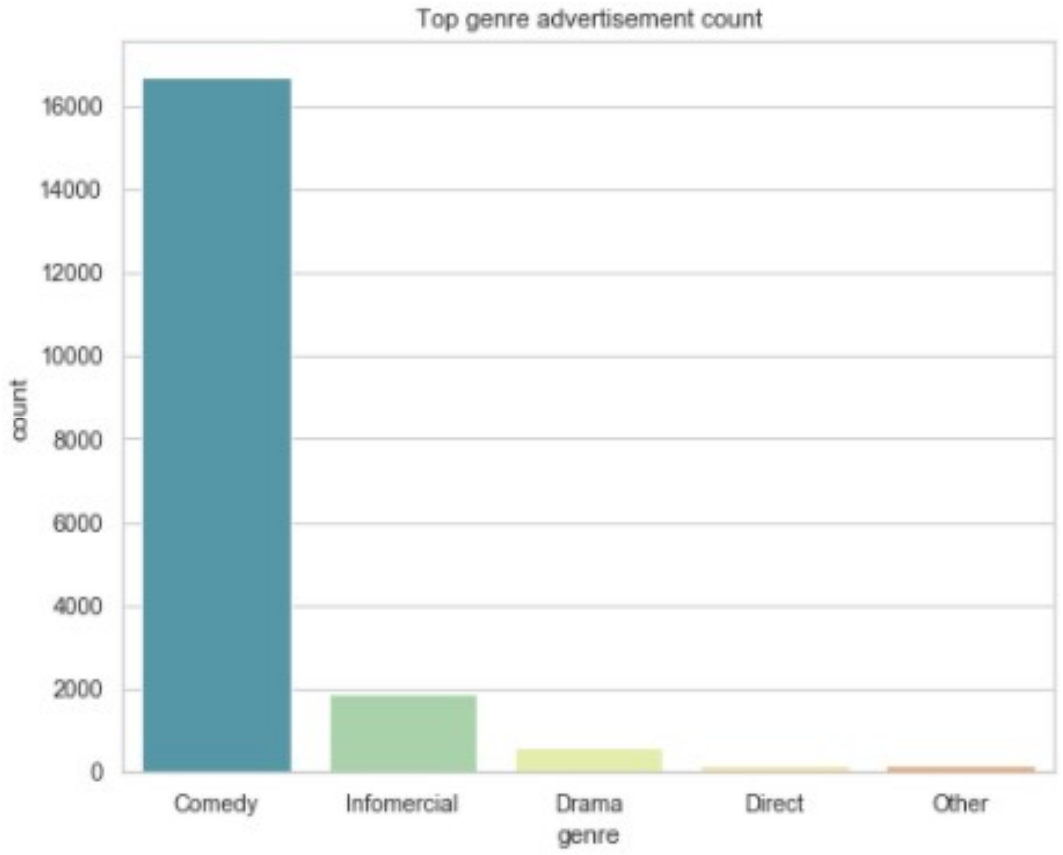
- Male add more value to net gain as compare to female.



EDA

Top Genre wise Advertisement Count

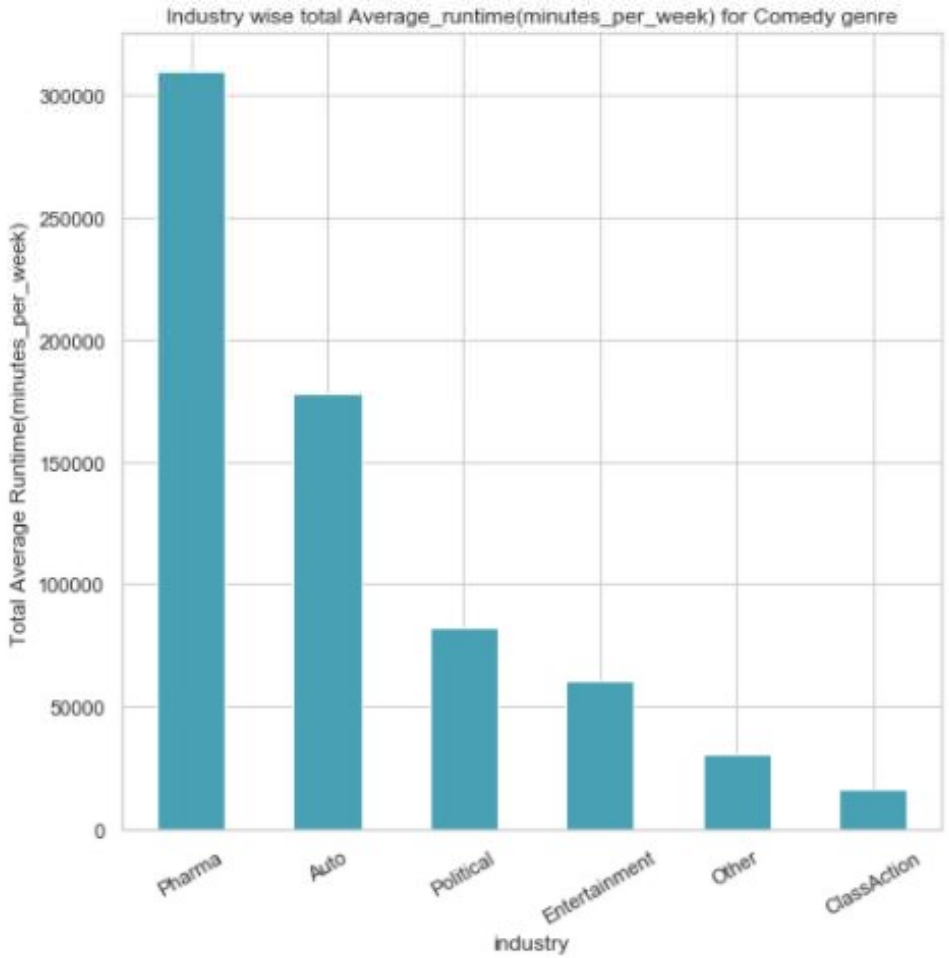
- Comedy is the most preferable genre.
- Direct is the least preferable genre.



EDA

Top Industries in Comedy Genre

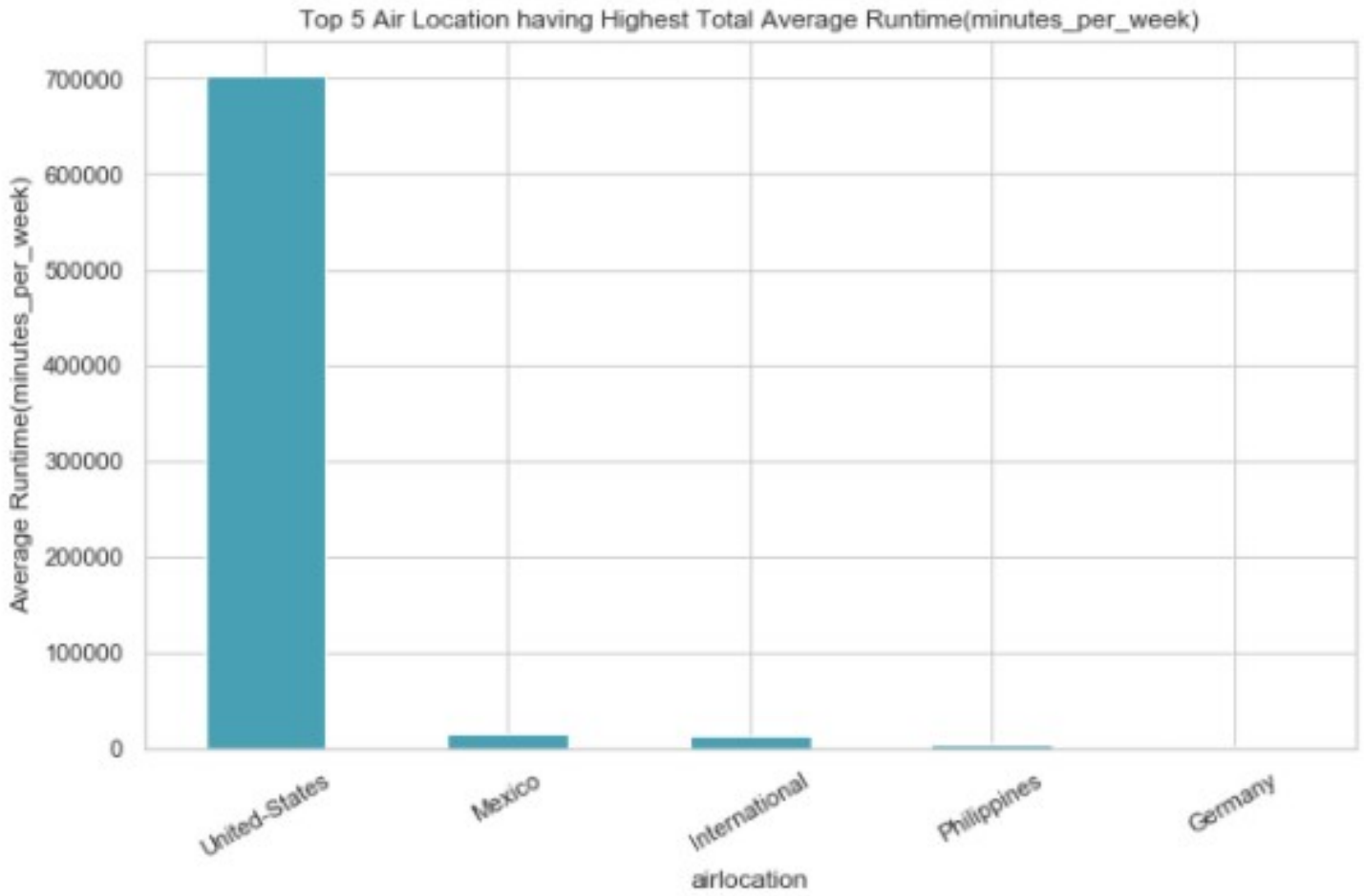
- Pharma industry having highest total average runtime for Comedy genre.
- ClassAction industry having lowest total average runtime for Comedy genre.



EDA

Top 5 Air Location for Avg Runtime

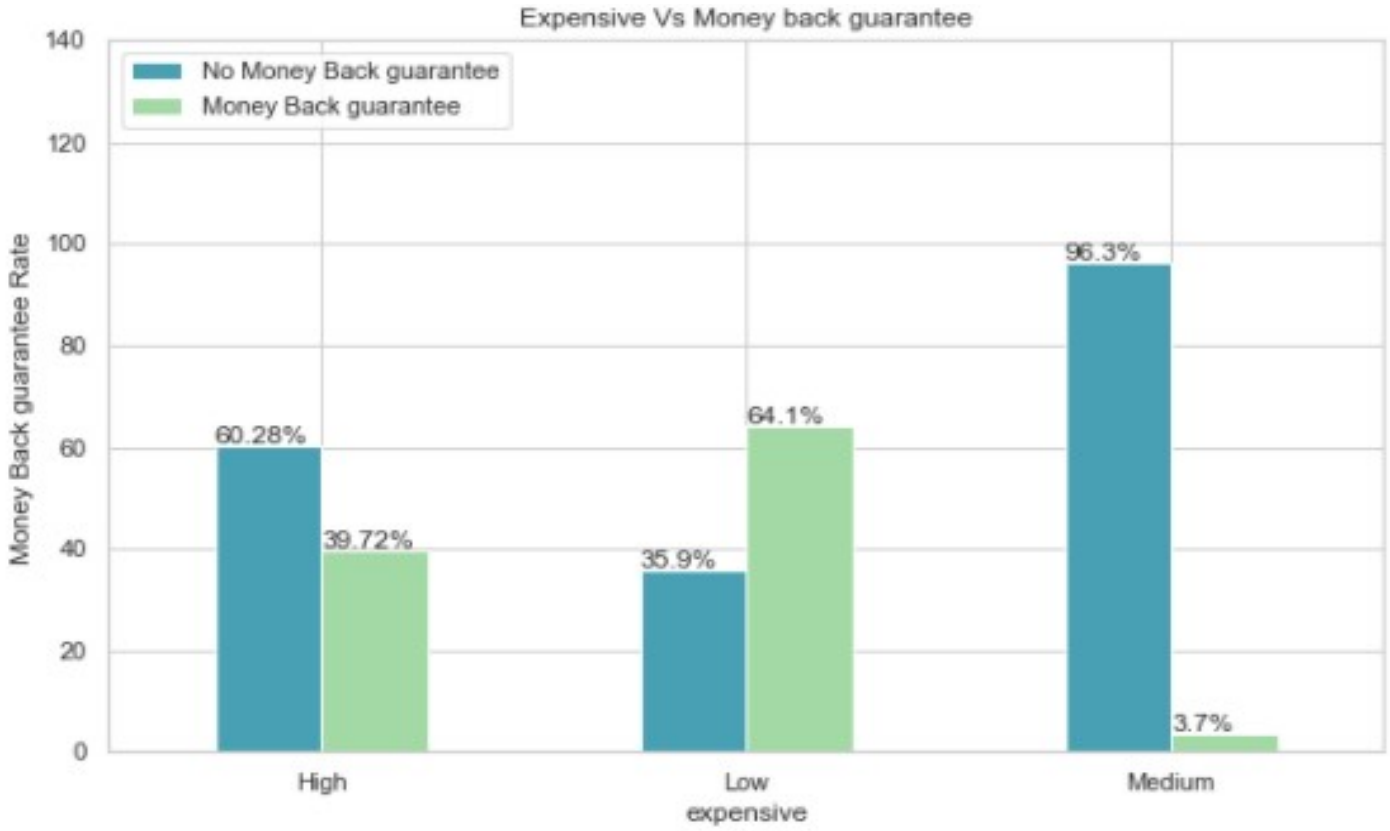
- United-States having highest total average runtime(minutes_per_week).



EDA

Expensive Vs Money back guarantee rate

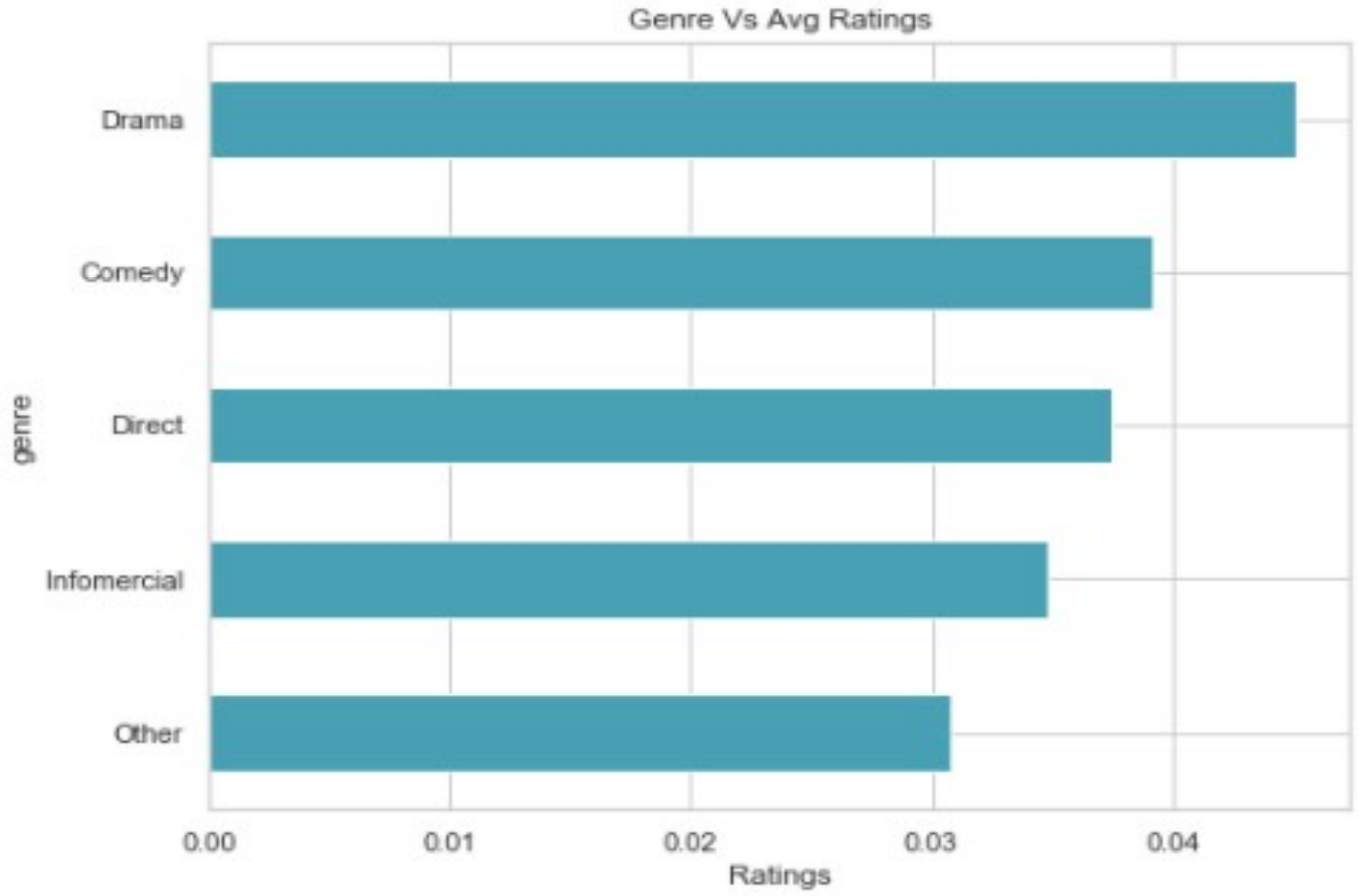
- High and medium expensive advertisements having low money back guarantee rate.
- Low expensive advertisements having high money back guarantee rate.



EDA

Genre Vs Average Ratings

- Drama advertisements having highest average ratings.



Pipeline

Missing Values :

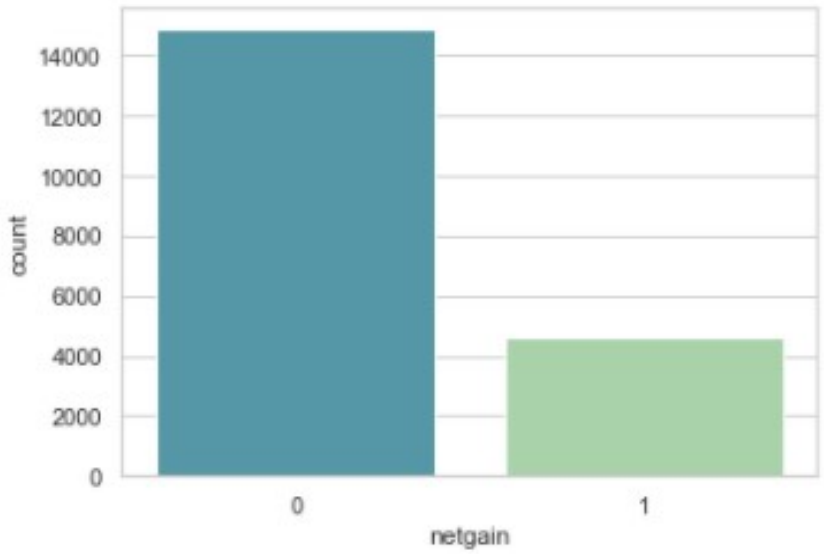
There are no missing values present in the data.

Null Values :

There are no null values present in the data.

Imbalanced Class :

0	1
14886	4650



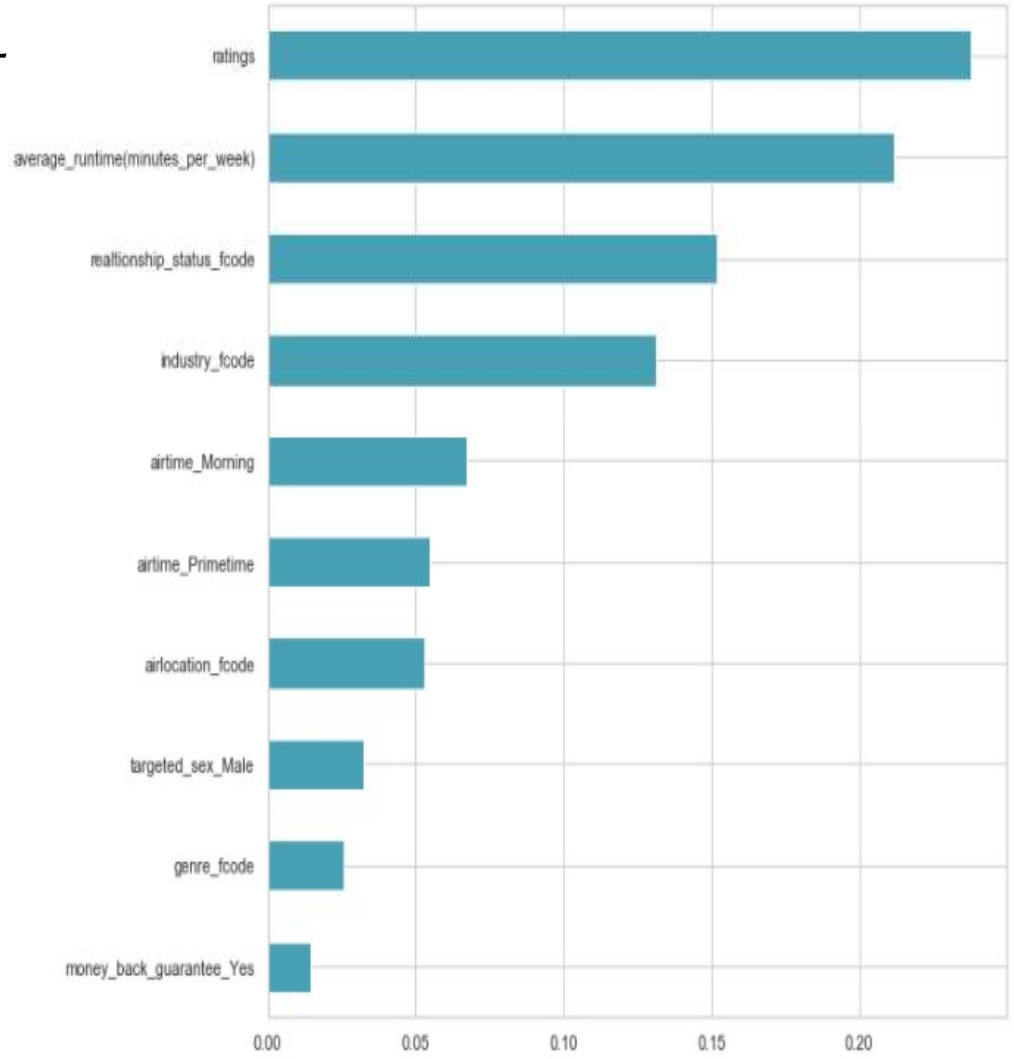
Feature Engineering

1. One Hot Encoding
2. Frequency Encoding

Column Name	Technique
targeted_sex	One Hot Encoding
airtime	One Hot Encoding
expensive	One Hot Encoding
money_back_guarantee	One Hot Encoding
realtionship_status	Frequency Encoding
industry	Frequency Encoding
genre	Frequency Encoding
airlocation	Frequency Encoding

Feature Selection

- Numerical Data(Co-relation)
- Categorical Data(Chi Square)
- Recursive Feature Elimination
- Extra Trees Classifier for feature importance
- XGBoost Classifier with Grid search



Model Selection

- Logistic Regression
- Decision Tree Classifier
- Random Forest Classifier
- XGBoost Classifier

Different Model Scores

- With Imbalanced Data

Model	F1 Score
Logistic Regression	0.40
Decision Tree Classifier	0.49
Random Forest Classifier	0.52
XGBoost Classifier	0.53

Balanced Data & Hyper parameter Tuning

Imbalanced data handling techniques:

- SMOTE
- SMOTE Tomek

Hyper parameter tuning techniques:

- Grid Search CV
- Random Search CV

Different Model Scores

- With Balanced Data

Model	F1 Score(SMOTE)	F1 Score(SMOTE Tomek)
Logistic Regression	0.77	0.77
Decision Tree Classifier	0.82	0.83
Random Forest Classifier	0.83	0.84
XGBoost Classifier	0.84	0.85

Final Confusion Matrix

Top two algorithms:

1. Random Forest Classifier : 0.84
2. XGBoost Classifier : 0.85

- From the above observations, it can be inferred that best performing model is XGBoost Classifier with F1 score of 0.85

		Actual	
		No Net Gain	Net Gain
Predicted	No Net Gain	3426	980
	Net Gain	429	3968

Recommendations

- There are very less net gain found for female targeted gender, in such case company should have to plan some future strategies to target this part of advertisement.

- As the low expensive advertisements provide maximum number of money back guarantee, so company can increase the cost of such advertisement slightly to earn more profit for such cases.

- As Comedy genre having maximum number of advertisement, so for this case company should have work on resources to improve the ratings of such advertisements.

Thank You!!!