

# Credit Card Spend ML Prediction Model

---

Presented By: Rabia ariq

Data Provided By: [kaggle.com](https://www.kaggle.com)

Springboard School of Data - 8/29/21

# Problem

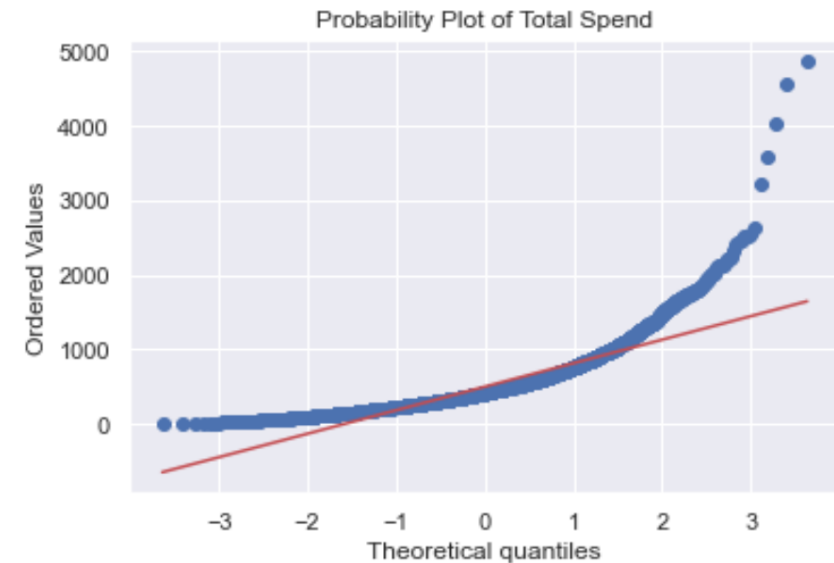
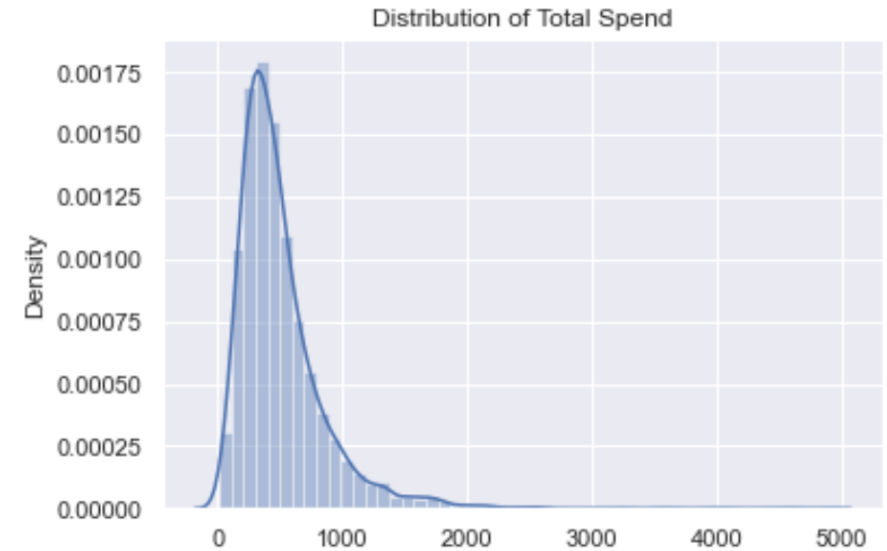
---

- Credit cards are great, convenient method of making payments, easier to carry than cash, helps you build a credit rating and are there if you need emergency cash. Lenders like to understand the factors driving the credit card spend.
- How can we use the data to understand what is driving the total spend (Primary and Secondary cards) and given those factors predicting the limit of their credit card spend.



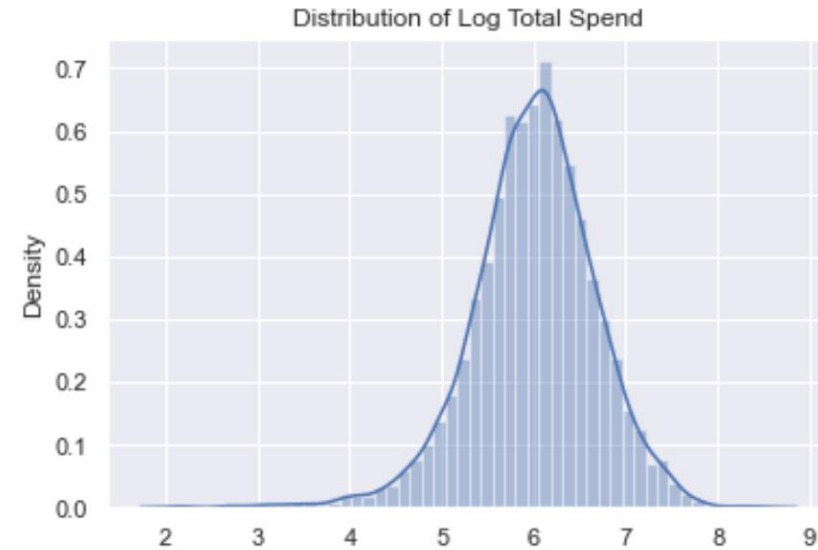
# Analysis

- Target variable is right skewed.
- The distribution of our target variable against the theoretical normal distribution shows that the target variable does not closely follow the normal distribution.



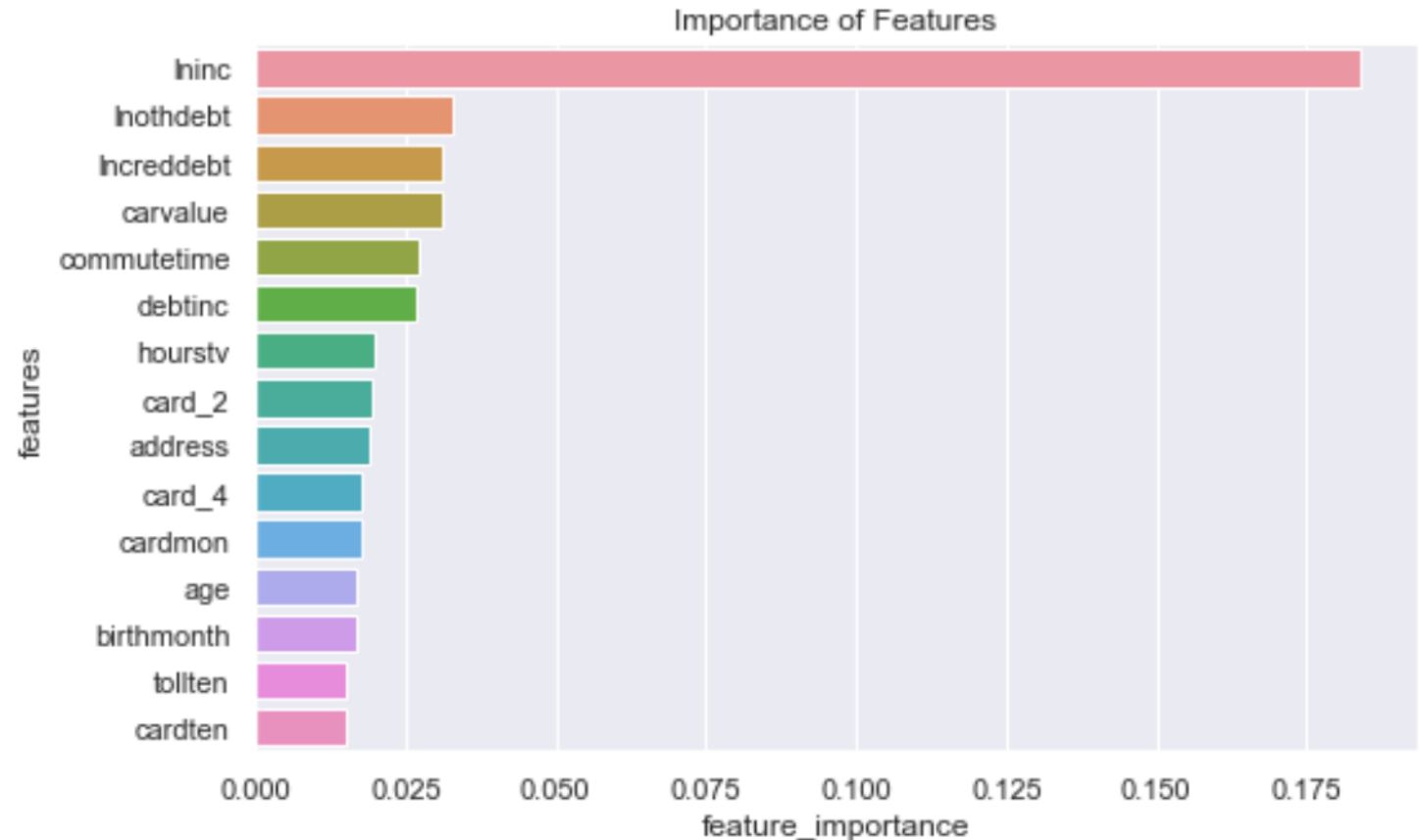
# Analysis

- Taking the log of the target variable greatly reduces the skewedness.
- The distribution of the log variable against the theoretical normal distribution shows to be a lot closer to the theoretical normal distribution.
- To make the modeling easier, the target variable is chosen to be the log of total spend.



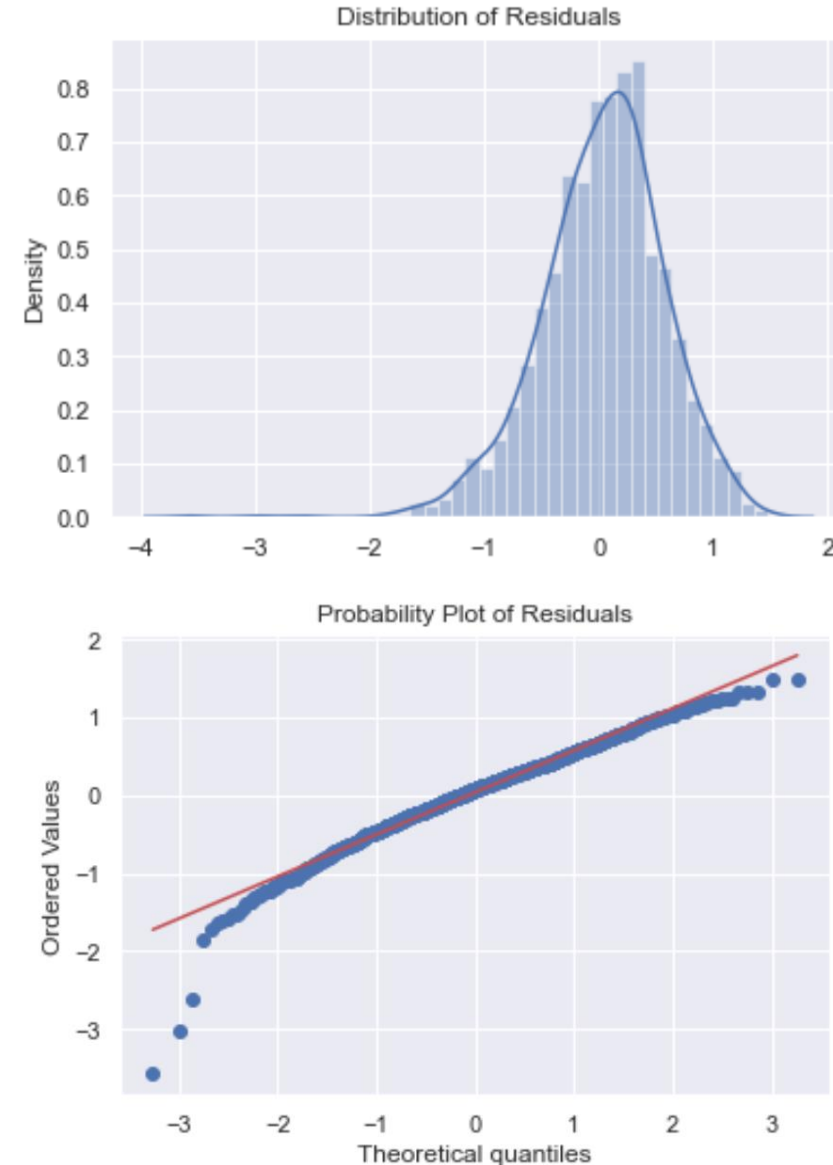
# Analysis

- The most important features for the model to make predictions were found by using Random Forest Regressor.
- The top four features were lninc, lnothdebt, lncreddebt and carvalue.



# Models

- The models used to predict the monthly credit card spend of customers were Ordinary Least Squares (OLS), RandomForest and XGBoost.
- The distribution of the residuals showed normal distribution and closely resembled the theoretical normal distribution.



# Model Evaluation

- The three models were evaluated using the metrics: mean squared error, mean absolute error, root mean squared error.
- By looking at the MSE, MAE, RMSE, the Ordinary Least Squares model performed the best.
- From evaluation OLS is used for predictions of monthly total spend.

	MSE	MAE	RMSE
<b>OLS</b>	93467.195826	197.760448	305.724052
<b>RandomForest</b>	103425.755175	208.295764	321.598749
<b>XGBoost</b>	108212.096835	216.376562	328.956071

## Prediction

- The target variable was log of the total spend.
- To get the predictions of the total spend, took the inverse of the values predicted by the OLS model.

	Actual_total_spend	Pred_total_spend
3010	151.91	366.230224
1376	311.50	338.113679
4368	279.97	435.573564
713	251.89	324.737996
206	2425.02	877.573900



# Conclusion

---

- The work done in the analysis of the data is important for banks in determining the credit limit of the customers.
- Banks look at different factors in determining the credit limit of an individual, one of them being their spending history and spending habits. So, the banks can use the predictive model to help in determining the credit limit of the customers who open apply for a credit card in the future.