# DATA SCIENCE PORTFOLIO:

## Predicting User Churn to Strategically Retain Customers

Nick Liszewski

# Summary of Data

1. Sample data is retrieved from Kaggle, a popular data science website, to practice data analysis.

2. Data is from a credit card company, including user data, as well as whether the user stopped using the company's service or not.

# Data Preparation

1. Convert string fields to usable integers.
    - "Male" and "Female" transformed to 1 and 0 respectively
2. Drop observations with missing values (24% observations lost)
3. Normalize values with Z-score Normalization
    - Makes different variables comparable in Euclidean space.
4. Data is separated into "train" and "test" sets to validate accuracy. (80/20 split)

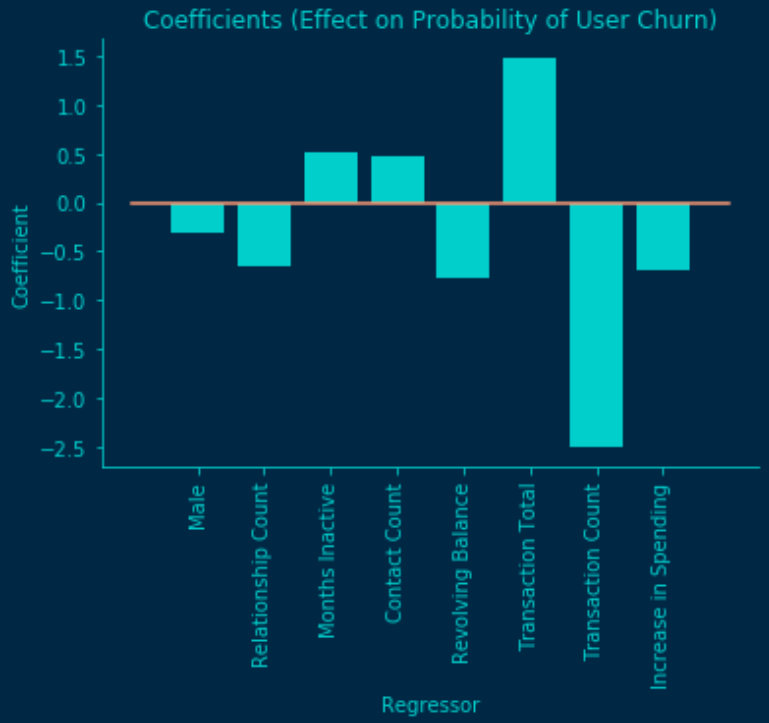
# The Problem

- Our company wants to predict which users are likely to stop using our services soon.
- How can our different departments intervene to retain these customers?
    - Marketing: Which demographic should we target in our advertising campaigns?
    - Sales: Which product does each demographic enjoy most?
    - Executives: Should we shift the business model to focus on the most loyal customer types?
    - Support: How should we amend our support system to better retain our customers?

# Logistic Regression: Which features co-occur with high user churn?

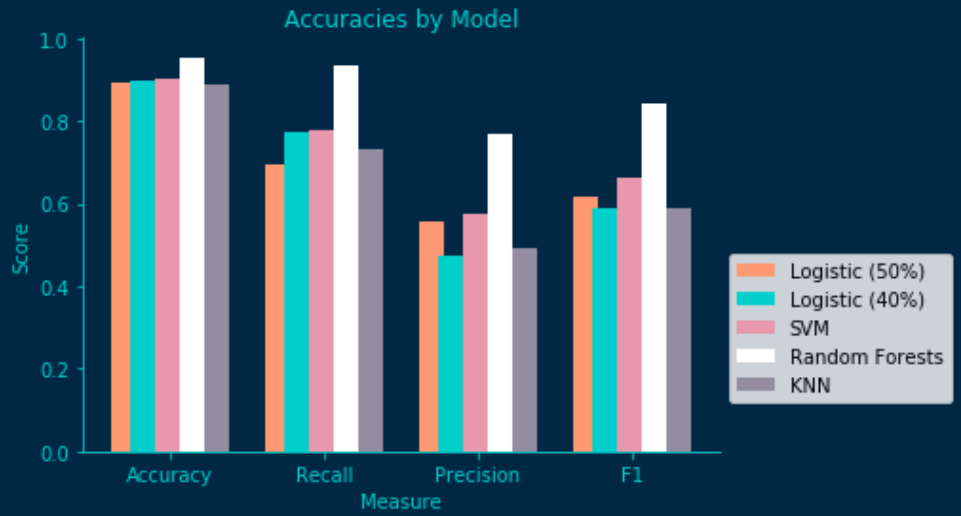
- We use Recursive Feature Elimination to quickly select the top 8 most impactful coefficients on user churn

1. Male users are less likely to churn
2. Users with more of our products are less likely to churn
3. Inactive users are more likely to churn
4. Users who contact us more frequently are more likely to churn
5. Users with a high revolving balance are less likely to churn
6. Users with high total transaction amounts are more likely to churn
7. Users with many transactions are less likely to churn
8. Users who show an increase in spending are less likely to churn



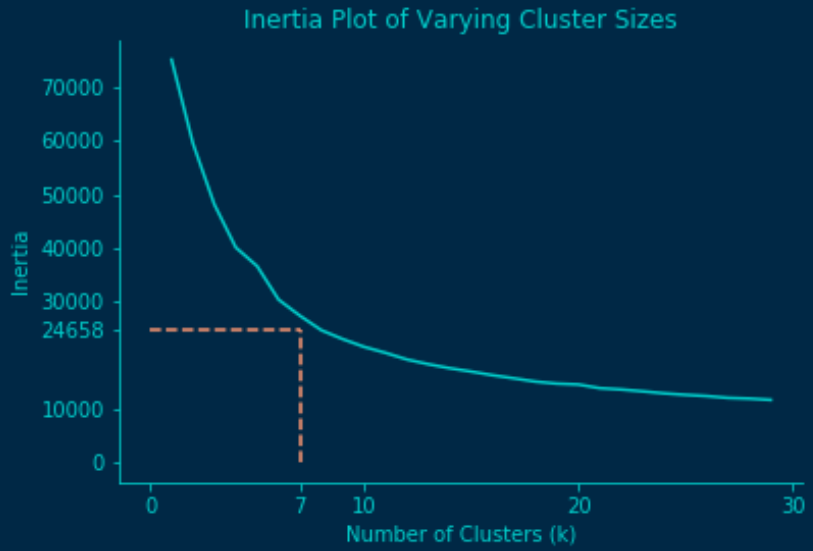
# Machine Learning: Predicting User Churn

- Each model takes input data about the customer
    - Gender, revolving credit utilization, etc.
- Based on this input data, each model predicts whether the user will churn or not.
- The Random Forests Model performs best in all measures of accuracy.
    - With this trained model, our company will be able to intervene in 95% of cases where a customer will soon stop using our service
    - The model has 77% precision, meaning that among those flagged, only 23% will continue using our service regardless of intervention.



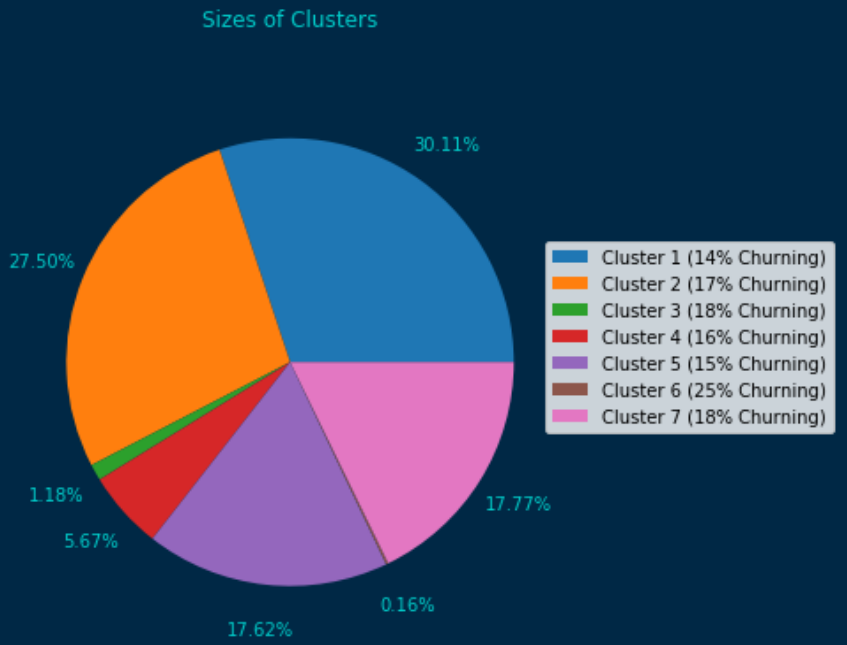
# Clustering Our Users: Optimal K

- We want to create clusters of similar users so we can see which types of people have the highest churn rates.
- How many clusters should we choose?
    - To answer this, we plot how much our cluster means move (inertia). We want a low number of stable clusters.
- At k=7, we see our inertia levels off, indicating a good cluster size to settle on.



# Cluster Demographics

- Cluster 5 is churning at 25%! But they are only 0.16% of our customer base.
- Meanwhile, cluster 1 represents 28% of our customer base, churning at 17%.
    - This cluster is mostly low-income, married women who don't have strong preferences for any one of our products.
    - How might our product development team better tailor products to retain this group of customers?
- We can apply this same clustered demographic analysis to each one of our clusters, to provide better recommendations to each department



# Conclusion

- Company departments should strategically alter their methods if our goal is to retain our existing customers.
- Our product development team can create a menu of products to meet the needs of each customer cluster
- Our marketing team can advertise while directing specific demographics to their best-fit products
- Our sales team can help guide certain demographics of customers to the products they will probabilistically be most satisfied by
- Our support team should be streamlined to minimize interactions with customers, while still solving their problems
- Executives should make decisions informed by which customer demographics are most loyal to the company
    - To preserve existing customer base
    - To expand into new customers whose needs aren't being met by our products