Telecom Customer Churn Analysis

Blake Tindol, Yesul Song, Sean Deery

IST-718 Big Data Analytics

Agenda

Customer Churn in Telecom

- Problem Background
- Data

Exploratory Analysis

- Demographic Data
- Customer Relationship Data

Models

- Logistic Regression
- Random Forest
- Support Vector Machine

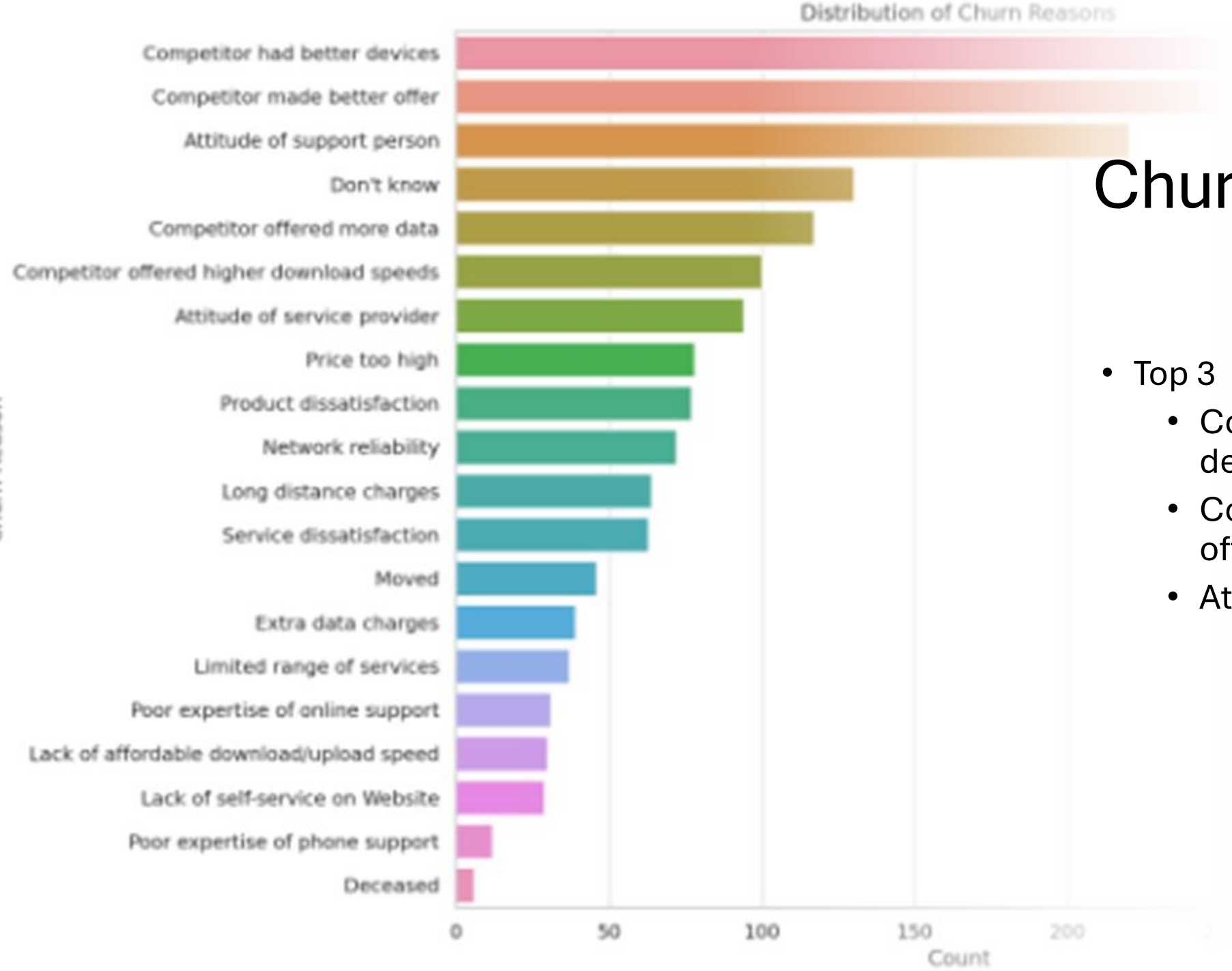
Customer Churn in Telecom

- Subscription Based Business:
  - Monthly,
  - 1-year,
  - 2-year contracts
- Revenue and long-term growth
- Data
  - Telecom Customer Churn Prediction dataset from Kaggle
  - Telecom Company in California
  - Customer Data Q2 2022
  - Population by zip code
  - https://www.kaggle.com/datasets/shilongzhuang/telecom-customer-churn-by-maven-analytics/data

Exploratory Data Analysis

Distributions

Correlations with Churn

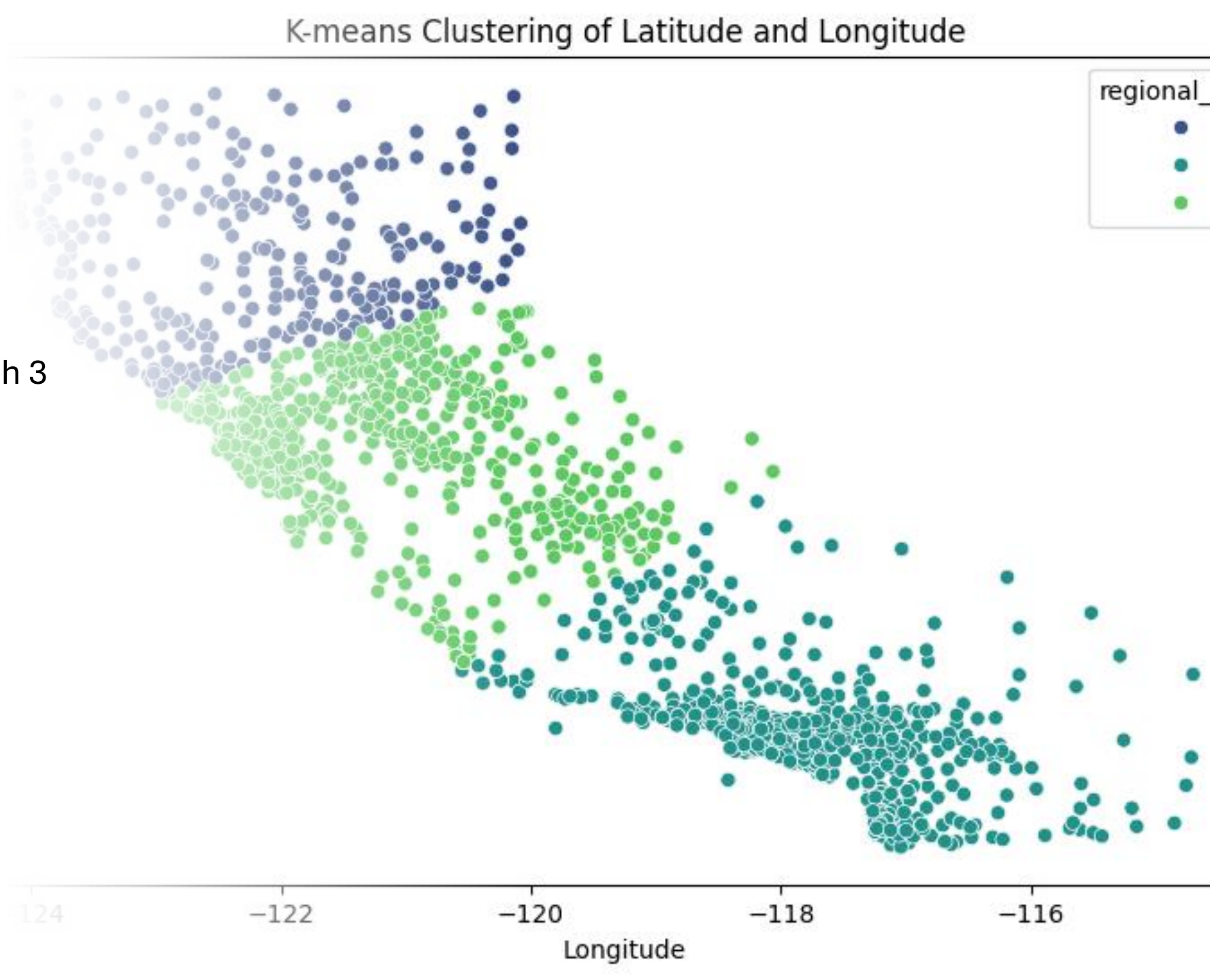


Churn Reason

- Top 3
  - Competitor has better devices
  - Competitor made better offer
  - Attitude of support person

Location

- K-Means Clustering with 3 clusters
  - Northern CA
  - Central CA
  - Southern CA



Predictive Models

Logistic Regression

Random Forest Classifier

Support Vector Machine

Models Test Accuracy

Logistic Regression

- Unbalanced Data: 83.7%
- Down-sampled Data: 78.1%
- Up-sampled Data: 78.9%

Random Forest

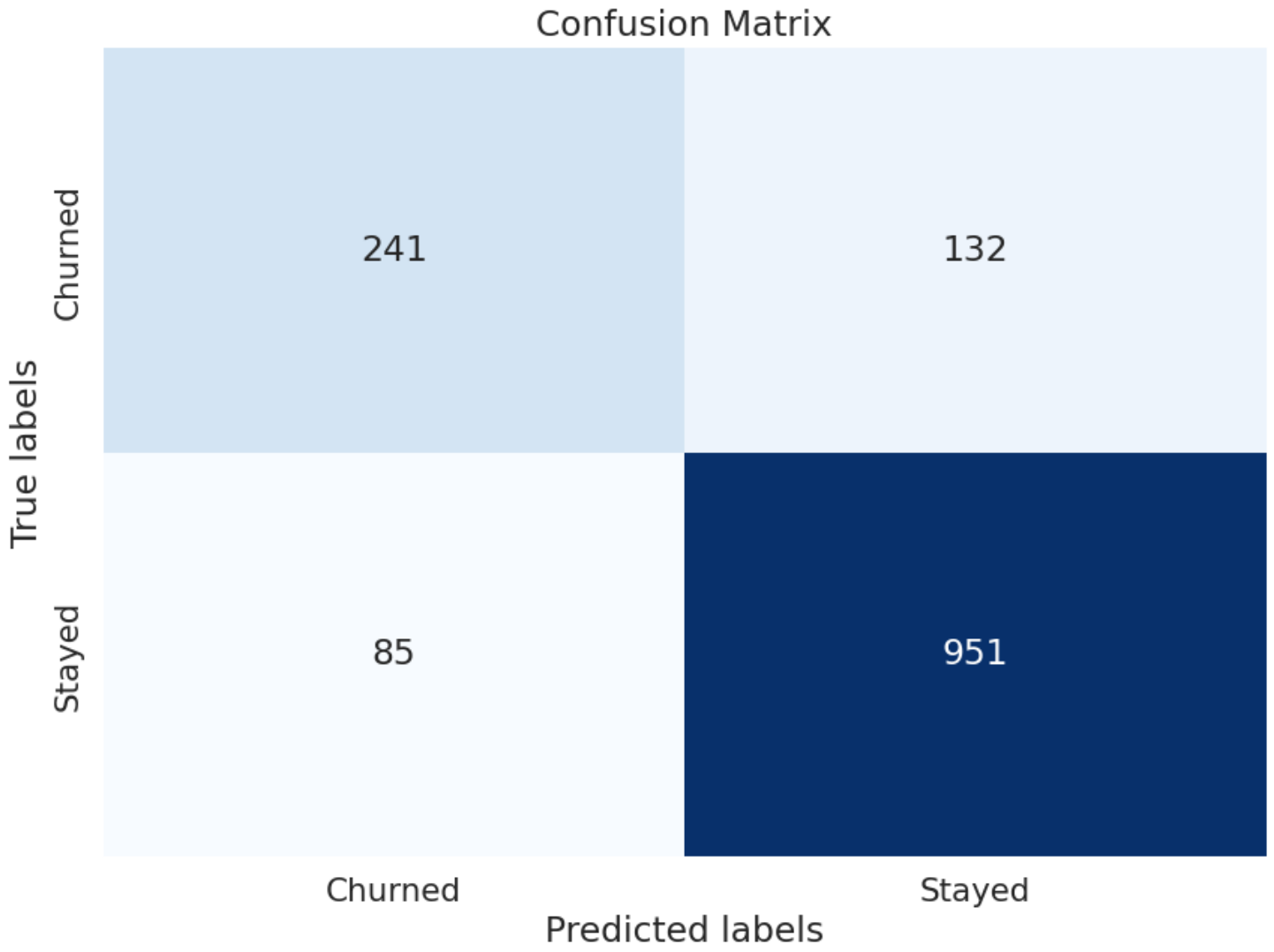
- Unbalanced Data: 84.6%
- Down-sampled Data: 79.2%
- Up-sampled Data: 81.5%

Support Vector Machine

- Unbalanced Data: 84.6%
- Down-sampled Data: 77.1%
- Up-sampled Data: 76.3%

Best Model

- Support Vector Machine: 84.6% Test Accuracy
- Unbalanced Data



Final Thoughts

- Best model was achieved with Support Vector Machine
  - 241 of the 373 churned customers in the test set were predicted as churned
- Recommendations:
  - Pricing Strategy Reevaluation
  - Enhancing Customer Engagement
  - Geographical and Demographic Targeting
  - Service Improvement