TELECOM CUSTOMER CHURN PREDICTIVE SOLUTIONS

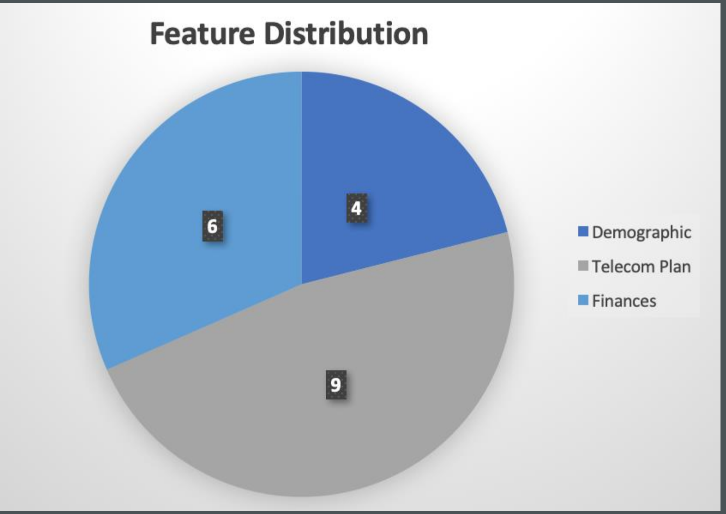
How to Limit the Churn Rate:
The definitive plan for limiting churn

RELATING

- **Goal –** Better understand churn behavior in order to retain customers
  - **Benefits –** Increase *profit margin* and improve *business reputation*
- **How** – Explore multiple *Classification* models to predict customers who churn vs. customers who do not churn
  - **Measure** – In order to limit data set bias, we will use *AUC* as our metric to determine how well our model performs
- **Success** – A definitive metric that results in >.60 AUC with a negligible turnaround time

EXPLAINING

- 5282 records, 19 features, 1 target variable
- ~26% of customers churned
- Average Monthly Charge
    - Not Churn: $61.27
    - Churn: $73.52
    - Competition: $50-$85



- Majority Characteristics of Churn Customers
    - Gender neutral
    - Not Senior Citizens
    - No partner/dependents
    - Phone Service
    - Online Security
    - Tech Support
    - **Month-to-Month contract**

ADAPTING

- Data Cleaning:
  - Converted categorical variables to '0' and '1' (and '2' if necessary) integer representations
  - Converted numerical variables to float and replaced numerical null values with '0'
  - Scaled the data so that the numerical variables wouldn't overpower categorical ones
- Data Selection:
  - Explored benefits of using all variables and previously mentioned influential features
  - Best result: 'Contract','PaperlessBilling','PhoneService','gender', 'SeniorCitizen', 'Partner', 'Dependents','tenure','MonthlyCharges'

SELECTING

- Model Approach: Compare performance of four well known classification algorithms
  - Logistic Regression
  - Random Forest
  - Gradient Boost
  - Neural Network
- All Features and Specific Features
- Split Training Data
- Hyperparameter tuning
- Cross Validation

OUTLINING

- **Top Performer** – Logistic Regression using specific influential features and 60/40 train/test split
- **Success** – A definitive metric that results in >.60 AUC with a negligible turnaround time
  - **Results** – 0.71373 AUC which is 18% better performance than we were expecting

	Retain Customers (T)	Losing Customers (T)
Retain Customers (P)	1708 x $61 = +$104,188	213 x $73 = -$15,549
Losing Customers (P)	342 x $61 = +$20,862	378 x $73 = -$27,594

NAVIGATING

Because this schema is month-to-month, real-time prediction is unnecessary.

Deploy via **AWS Sagemaker** which supports Python and scikit-learn

Allows us to retrain monthly and re-deploy to production	Easy connection to BI tools for monthly reports