Credit One

Examination of Customer Defaults

Proposing BADIR Framework for analysis

BADIR

- Business questions
- Analysis plan
- Data collection
- Insights
- Recommendations

Why BADIR?

- Concise process merges business goals with data driven recommendations
- Five-step method starts with clear understanding of business questions
- Efficient data analysis remains focused on business questions
- Actionable insights achieved through sound data science structure

Business Questions

- What are the business questions?
  - Which customer features are related to defaulting and not defaulting on loans?
  - Can a model be built to better predict how much credit to extend to reduce loan defaults?
- What is the intent of the underlying the questions?
  - The problem is that there has been an increase in customers defaulting on loans in the past year
  - Credit One risks losing business customers if problem is not solved
  - The impacted segment are customers who are defaulting (what do they have in common?)
- What are stakeholders' current thoughts on the underlying reasons?
  - To be determined

Business Questions

- What business considerations are likely to impact the analysis?
  - Stakeholders: Credit One executives; Guido Rossum, Senior Data Scientist; determine others
  - Timeline to answer business questions: As quickly as possible, business critical
  - Cost: Business clients could be lost the longer problem is unsolved
  - Actions:
    - Investigate common features of customers who, a) default and b) don't default on loans
    - Build predictive models based on those features to better predict credit limits to extend
    - Answer question, "Can we predict credit limits with higher level of certainty of customer not defaulting?"

Analysis Plan

- What are the analysis goals?
  - Identify differences in customer features by default vs. not default
  - Build models to predict credit limits based on most impactful features to reduce likelihood of defaults
- What hypotheses are to be tested?
  - There will be key differences between customers who default on loans vs. those who do not
  - Credit limits can be predicted with improved level of certainty of customers not defaulting

Analysis Plan

- What data is required/available to test the hypotheses?
  - Six-month historical data of 30,000 credit card clients is available
  - Includes customer demographics, monthly billing/payment information, and default status
- What methodologies should be employed?
  - **Exploratory data analysis** (EDA) to determine relevant datapoints most useful for modeling
    - Exhaustive conditioning and exploration
    - Identify patterns and relationships
  - **Predictive modeling** on multiple model types, cycle of:
    - Optimizing models to minimize error and maximize accuracy
    - Evaluating models to determine if accurate enough to meet stakeholder needs

Analysis Plan

- What is project plan?
  - Data Science lead will initiate data analysis immediately on April 24, 2020
  - Timeline to complete business goals will be May 23, 2020:
    - May 9, 2020: Complete extensive exploratory data analysis, determine correlations
    - May 23, 2020: Complete report on confidence of predictive modeling results for executive team
    - Weekly updates/questions to Senior Data Scientist to ensure project is on the right track
  - Prioritization will be given to Customer Default project due to critical nature of problem
    - Ongoing projects will be reprioritized and communicated to pertinent stakeholders

Data Collection

- From where can the data be obtained?
  - From historical records of Credit One database of customers
- How must the data be cleansed and validated?
  - Check for missing data and handle appropriately (may need to insert averages, etc.)
  - Check unique values for each column and ensure they make sense
  - Change variable names to make more concise/understandable
  - Change variables with word values to number values so analysis is possible
  - If needed, remove unnecessary rows or columns not relevant for data analysis

Data Insights

Initial data insights:

- 30,000 total customers
- Includes prior 6 months of customer info
- 25 total variables
- Customer demographics include sex, education, marital status, age
  - Sex, education, and marital status are word values and must be changed to numeric values for data analysis
- No missing or duplicate data

To be determined in analysis process:

- What patterns are seen in the data?
- Are each of the hypotheses proven or disproven?
- How much confidence should stakeholders place in the results and model predictions?
- How do you rank your findings in terms of quantified impact on business?

Recommendations

Effectively present results with simple outline:

- Objective
- Background
- Scope
- Approach
- Recommendations
- Key insights with impact
- Next steps

Flowchart

- Business question
- Analysis plan
- Data collection
- Insights
- Recommendations

Potential pitfalls (and Solutions)

- Business goals are unclear (ask questions in the beginning to clarify)
- Unable to answer business questions (recommend additional data or modify questions to address business goals)
- Data quality not good enough (recommend additional data)
- Recommendations are vague (recommendations should be actionable and have positive impact)