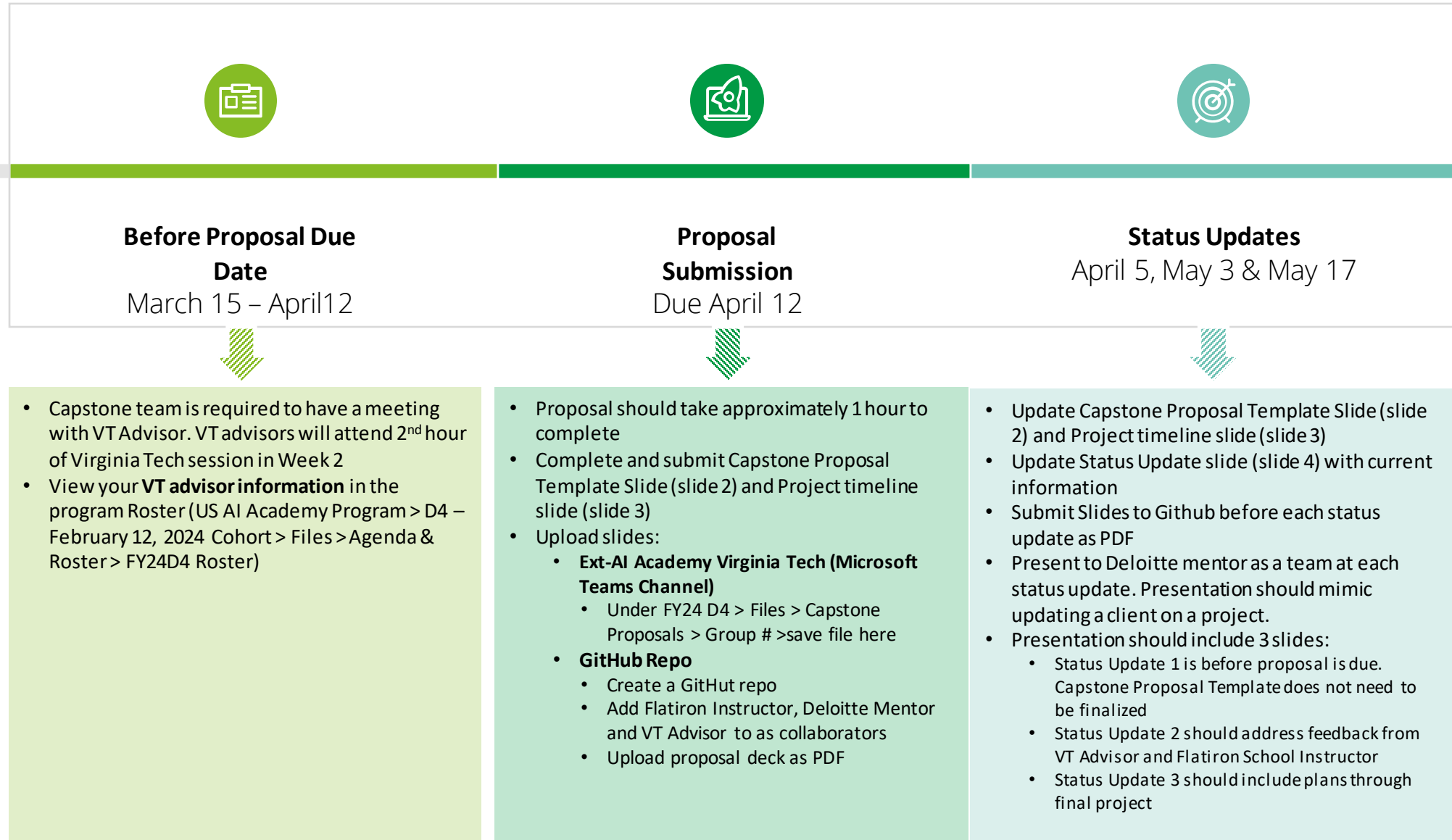


# Instructions on expectations for proposal and mentor status updates



# CAPSTONE PROPOSAL

Project Overview	
Project Name	LoyalLock: Predictive Modelling for Customer Segmentation and Churn Reduction

# CAPSTONE PROPOSAL

Project Overview	
Business Understanding*	<p><i>What problem are you trying to solve? What question are you trying to answer?</i></p> <p>In the competitive realm of credit card companies, each company offers a wide range of incentives to obtain a customer and keep them loyal to their services. From offering a reward system that allows customers to use their rewards towards travel, card statements, or gift cards to offering a low percentage of interest rates on monthly statements. Customer retention is a crucial factor for sustaining revenue growth and fostering long-term clientele and profitability.</p> <p>The key stakeholders include the credit card companies who seek to decrease customer churn, the customer service and marketing teams who may need to adjust their operations and strategies based on our insights, and the customers who will benefit from improved services and retention incentives.</p> <p>This proposal aims to recognize the importance of retaining clientele by implementing and developing a predictive analytics tool to forecast customer churn by enabling proactive measures to mitigate churning rates effectively. Our primary goal is to identify potential churning among our customers and harness the retention. We would investigate the identifying factors within the data set that we feel contribute to the customer churn. We would then perform evaluations with models to increase customer retention. By carefully analyzing these predictions and creating an analysis, credit card companies can then tailor their retention rates by implementing strategies to incentivize their current customers that are at risk of churning.</p> <p>To enable our predictive analytic models, we would require specific customer data. Using this data, we will be able to strategically plan and allocate resources to overcome potential challenges. The predictive modeling insights will guide changes to increase customer retention. This process may involve modifying current marketing strategies or enhancing customer service operations. However, with the correct approach and by leveraging the insights obtained from our predictive analytics, we are confident that any arising challenges can be effectively addressed, leading to improved customer retention and business growth.</p>

# CAPSTONE PROPOSAL

Project Overview	
Client/Project Motivation *	<p><i>Who is your client? What is the motivation behind your project?</i></p> <p>The client for this proposed solution would be credit card companies that are seeking ways to enhance customer retention strategies and reduce their churning rates to sustain revenue growth, clientele, and profitability. The motivation for this proposal falls under the umbrella of revenue protection, customer satisfaction, and competitive advantage while maintaining clientele retention.</p> <p>Revenue protection is vital when it comes to credit card companies. Credit card companies rely heavily on customer retention for their sustained growth. Therefore, customer churn results in a loss of revenue for the company. By looking retrospectively at the data collection and implementing predictive analytics to forecast customer churn and foresee possible issues, companies can proactively intervene with targeted retention strategies which in turn results in revenue streams and improving profitability.</p> <p>Customer satisfaction is also an important pillar when preventing customer churn. Retaining customers is not only about revenue growth and financial gain but also about fostering long-term relationships and customer loyalty. This in hance would benefit the customer and the company as it would lead to building a trustworthy relationship with the client and hopefully spreading awareness to other clients looking to sign up for credit cards thus expanding customer numbers.</p> <p>In addition to revenue growth and fostering customer satisfaction, another key motive for this proposal is competitive advantage. Leveraging predictive analytics for churn reduction provides a competitive edge through targeted marketing and superior customer service. The client will be driven by the goal of enhancing operational efficiency. Predictive analytics can help the client make informed decisions in response to the insights provided. The client will stay informed about business strategies, which helps when identifying threats and opportunities in advance.</p> <p>The client will be driven to set industry standards. The client will find great motivation in leveraging predictive analytics to anticipate market changes and evolving the needs of their customers. This approach will enable the client to stay ahead of industry trends, set new benchmarks for customer satisfaction and challenge new market opportunities which will strengthen their competitive edge.</p> <p>Regarding a project perspective, a key motivating factor would be to build an adaptable and scalable model. The predictive model should be capable of managing data complexities if the business expands and/or if the data of the customer base is to be more diverse.</p>

# CAPSTONE PROPOSAL

Project Overview	
Market / Industry	<p><i>What market does your solution serve?</i></p> <p>The dataset selected contains a wealth of customer information collected from within a consumer credit card portfolio, with the aim of helping analysts predict customer attrition. The market that our proposed solution aims to serve is credit card companies therefore falling into the finance industry.</p>

# CAPSTONE PROPOSAL

Project Overview	
State of the Art*	<p><i>How has this problem been approached in the past?</i></p> <p>Historically, customer retention has been approached by credit card companies through personalized customer service improvements and marketing. Using data analytics and AI to address customer retention is not a new approach. An increasing number of companies are adopting technologies to implement preventive measures. Many firms seeking to improve their customer retention will employ experts to construct machine learning models. Machine learning techniques and strategies have been used since 2004. It's important to note that the most effective strategy to prevent customer churn is to gain a comprehensive understanding of the customers, which can be achieved by analyzing both historical and recent customer data.</p> <p>In the past credit card churn has also been analyzed using customer surveys. Customer surveys gathered feedback directly from customers wanting to participate. The feedback was used to understand customer satisfaction and dissatisfaction which helped provide insights to churn.</p> <p>Furthermore, credit card companies have engaged in analytical studies which have included competitive, financial, and behavioral analysis. These firms have examined customers' interactions with their credit cards, aspects such as purchasing patterns, payment behavior, balances, and interest rate usage. Detailed investigation helps in understanding the likelihood of customer churn.</p> <p>While these traditional approaches may have been successful, credit card companies fail to address further problems related to loyalty and customer satisfaction.</p>

# CAPSTONE PROPOSAL

Project Overview	
Success Metrics (Evaluation)*	<p><i>What metrics will you use to determine success of your business recommendation?</i></p> <p>The success of our business recommendations will be assessed using a comprehensive set of metrics targeted at evaluating various aspects of customer retention. The primary metric for evaluating the success of the predictive analytics and retention strategies will be churn rate. Churn rate measures the percentage in which customers decide to discontinue their credit card services with a company. A decrease in churning rates will demonstrate the effectiveness of the implemented proposed solutions in retaining customers from discontinuing their relationship with said credit card company.</p>

# CAPSTONE PROPOSAL

Project Overview	
Scalability*	<p><i>If successful, would it be possible to scale the project to size required to meet your client's needs? What challenges do you anticipate?</i></p> <p>If successful, the predictive model should be able to manage data complexities to the size required to meet the client's needs, particularly with the help of technologies and digital tools. However, our team is aware that our client may face challenges. Technically, as the scale of the project expands, there may be a need for more advanced tools for data processing and analytics. The challenge resides in affording and implementing the tools that effectively manage an increase in data complexity. Additionally, an increase in data complexity also means an increase in the data that must be properly managed. The client needs to ensure that the customer data is following privacy laws and regulations to protect individuals' privacy rights and ensure that they comply with current data privacy and security standards.</p> <p>In terms of keeping up with competitors' reward offerings, it is important for the client to understand and create a realistic balance between offering competitive rewards to retain customers while also staying true to their brand identity. The client may struggle to match or better the competitive advantage of other companies in terms of those reward offerings. However, through the scaling process, the client must not neglect its core business values and brand identity.</p> <p>Finally, when it comes to scaling up reward offerings and promotions, it may not be practical to offer the same level of rewards to every client. The client must understand that maintaining such a level may not be feasible as the scale increases.</p>



# CAPSTONE PROPOSAL

## Modeling Requirements

Data Type (e.g., numerical, categorical)

What type of data is needed to answer the posed question(s)? See tables below

Column name	Description	Data type
CLIENTNUM	Unique identifier for each customer.	Integer
Attrition_Flag	Flag indicating whether or not the customer has churned out.	Boolean
Customer_Age	Age of customer.	Integer
Gender	Gender of customer.	String
Dependent_count	Number of dependents that customer has.	Integer
Education_Level	Education level of customer.	String
Marital_Status	Marital status of customer.	String
Income_Category	Income category of customer.	String
Card_Category	Type of card held by customer.	String
Months_on_book	How long customer has been on the books.	Integer
Total_Relationship_Count	Total number of relationships customer has with the credit card provider.	Integer
Months_Inactive_12_Mon	Number of months customer has been inactive in the last twelve months.	Integer
Contacts_Count_12_Mon	Number of contacts customer has had in the last twelve months.	Integer
Credit_Limit	Credit limit of customer.	Integer
Total_Revolving_Bal	Total revolving balance of customer.	Integer
Avg_Open_To_Buy	Average open to buy ratio of customer.	Integer
Total_Amt_Chng_Q4_Q1	Total amount changed from quarter 4 to quarter 1.	Integer

Total_Trans_Amt	Total transaction amount.	Integer
Total_Trans_Ct	Total transaction count.	Integer
Total_Ct_Chng_Q4_Q1	Total count changed from quarter 4 to quarter 1.	Integer
Avg_Utilization_Ratio	Average utilization ratio of customer.	Integer
Naive_Bayes_Classifier_Attrition_Flag_Card_Category_Contacts_Count_12_mon_Dependent_count_Education_Level_Months_Inactive_12_mon_1	Naive Bayes classifier for predicting whether or not someone will churn based on characteristics such	String

# CAPSTONE PROPOSAL

Modeling Requirements	
Data Source	<p>The source of dataset selected is from Kaggle, Kaggle is defined as the world’s largest data science community on internet. In this platform you can found datasets, use notebooks for running code and enlist to competitions and solving data science challenges.</p> <p>The Kaggle page grabbed its data from Zenodo.org where all the credit card information was stored.</p> <p>Access: <a href="#">Prediction of Churning Credit Card Customers (zenodo.org)</a></p>

# CAPSTONE PROPOSAL

Modeling Requirements	
Data Preparation Steps*	<p><i>Which tools or methods are needed to prepare the data before using it?</i></p> <p>The dataset that we have has 20 features that consist of various strings and integers. We will need to ensure that all the data is in numerical form before beginning any sort of analysis on the data and before any training begins. Thus, we will have to implement a one-hot-encoding technique to create dummy variables for the various features that are string values.</p> <p>We will conduct extensive exploratory data analysis to determine if there are any features that are imbalanced one way or another (i.e., not providing much information for the prediction). The tools we will need to complete these steps are the one hot encoding library on python as well as pandas and matplotlib to perform our EDA. If we notice anything abnormal regarding the features, missing values, imbalances, or outliers, we will need to use additional tools to handle those cases such as SMOTE and feature selection (this is discussed more in the following section). In addition to the previously mentioned tasks, we must standardize all the features to ensure that the model interprets each feature on the same scale.</p>

# CAPSTONE PROPOSAL

Modeling Requirements	
Data Challenges*	<p><b>Missing data:</b> Observing the Kaggle information and analyzing the dataset, there is not missing values from any feature information, which is exceptional for the teamwork in data cleaning process. Most of the problem results could be on taking decisions about but this will not be the case.</p> <p><b>Class imbalance:</b> There are some features with this type of cases, talking about these problems, for categorical values this focus on what type of profile or category is better to analyze, the risk of not have this in consideration will result on a model that is not capable of approaching reality, since it will only consider the categories with the greatest amount of data. Best way to approach on this is make a configuration in hyperparameters to split the same amount of data by category for training the model, this will result on a balanced classification.</p> <p><b>Data with normal distribution:</b> Observing with detail the dataset features, this could be a minimal risk, because most of the integer data is statistically normal, this means that data will not be imbalanced in the case of this type of data. Finally, despite there are some outliers, is not representing more of 5% of the total data in that feature.</p>

# CAPSTONE PROPOSAL

Modeling Requirements	
Modeling Techniques*	<p>Despite there are many options of classification algorithms, the best options to approach this case will be selected from the simplest model to the most complex model, according to the implementation to this project:</p> <p><b>Logistic Regression:</b> In some cases, a complex model could not be the best way to approach a problem, and ever classification problem needs to start from the simplest way to get a base accuracy score. Another perspective is about the problem case, because the objective of this project is to determine between yes or not.</p> <p><b>K-Nearest Neighbor:</b> In this model, it supposed to classify a result based on what group of point is nearest positionally. This could be other model to use because there are a lot of features that can be selected.</p> <p><b>Decision Tree:</b> This model can be excellent for many classification problems; In some cases, this could be a base model cause is ever worth it to use. Around the decision rules, it could be made according to the features selected.</p> <p><b>Support Vector Machines (SVM):</b> According to different features, it is created a linear that separate two groups, despite is a good model to use in classification problems, is probably that data is not linearly separable.</p>

# CAPSTONE PROPOSAL

Modeling Requirements	
Target Variable*	<p><i>Indicate the target variable and connect the choice to the proposed problem.</i></p> <p>The column 'Attrition flag' will contain the values for us to try to predict with the model. This is because the data within this column corresponds to whether an individual customer churned out of the company's services or not. It is stored as a binary variable where 0 is an existing customer and 1 is a customer that left. Using this as our target will allow us to train the model with on all the data including age, income, education, etc. and later predict whether a customer will churn out or not based off these characteristics. This will allow us to preemptively create a group of individuals that are likely to leave the company soon based on their characteristics. Once we have obtained this group of people, we can begin to customize a business strategy to improve the probability that the customer does not churn out.</p>

# CAPSTONE PROPOSAL

Modeling Requirements	
Tools/Methodologies	<p><i>Indicate the tools you plan to use to solve the problem. Name specific Python libraries or particular techniques that will be used.</i></p> <ul style="list-style-type: none"><li>• Pandas</li><li>• Scikit-Learn</li><li>• Matplotlib</li><li>• Numpy</li><li>• Feature Selection</li><li>• Dimensionality Reduction</li><li>• Dummy Variables</li><li>• Oversampling of data</li></ul>
Regression or Classification problem	<p>Our use case is a classification problem.</p>

# Project Timeline

Recommended timeline – Adjust/customize based on your team plan. This should be updated as part of your proposal and status update

Week Number	Status Update	1	2	3	4	5	6	7	8	9	10	11	12
Milestones				Mentor Status Update	Proposal Due	VT Feedback Available	Flatiron Feedback Available	Mentor Status Update		Mentor Status Update		Practice Pres.	Final Pres.
Formulate business problem	Complete												
Find data set	Complete												
Perform literature review	On Track												
Solution hypothesis	On Track												
Implement solution	On Track												
Preliminary results	On Track												
Final Results	On Track												
Evaluate and Benchmark	On Track												
Presentation Prep	On Track												

Indicate:  
Complete  
On Track  
At Risk  
Off Track