Report 2

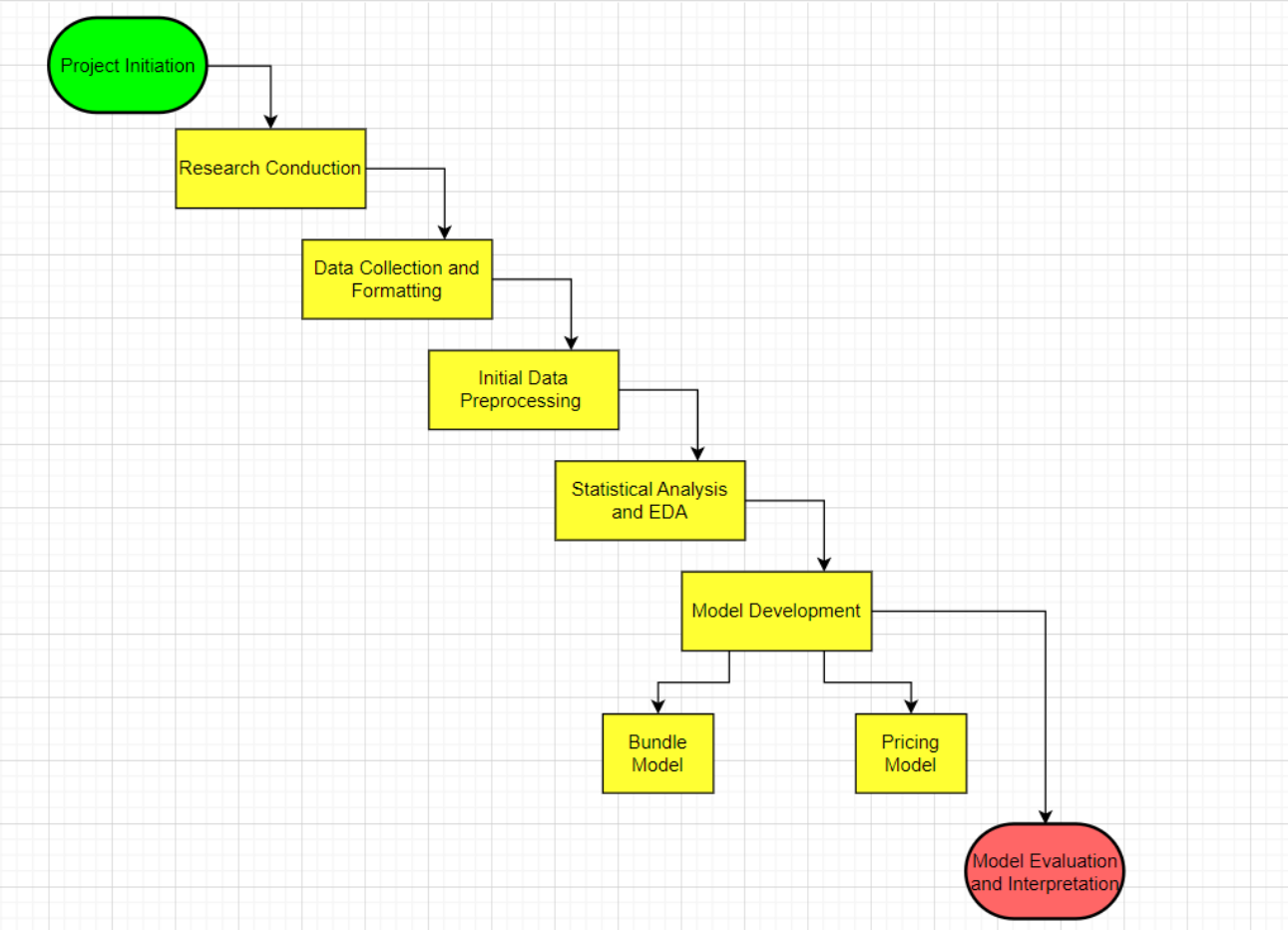
Name: **Musadig Aliyev**

Project Title: **Implementations of Product Bundling and Pricing Strategies using Machine Learning Techniques in the Banking Industry**

Selected Methodology:

My research topic can be considered a quantitative way of research because it involves the application of machine learning techniques to analyze and model data in order to derive insights and make predictions. Quantitative research typically relies on collecting and analyzing numerical data, which can be obtained from various sources such as databases, surveys, or historical records. In this case, I will collect relevant data related to product bundling and pricing model from bank's databases. In addition, the research specifically focuses on the implementation of machine learning techniques which can be classified as experimental research. Furthermore, I would use statistical analysis to analyze the collected data and derive meaningful insights. Finally, I would evaluate the effectiveness of different product bundling and pricing models and measure their impact on based on success criteria. Overall, this research topic involves the collection and analysis of numerical data, the application of machine learning techniques, and the utilization of statistical analysis to draw conclusions.

Project Plan Flow Diagram:



Project Timeline (11 weeks):

1. Week 1-2:

- Explore different areas of advanced analytics
- Establish the project's scope and objectives to determine its intended focus and desired outcomes.
- Conduct an extensive literature review on product bundling and pricing models

2. Week 3-4:

- Define concept of bundling and pricing model in banking
- Check data availability in bank for those models
- Get approval for use of data in academic research
- Create sample bundles
- Initialize evaluation methods and success criteria

3. Week 5-6:

- Raw transactional data Collection for Bundling Model
- Initial data collection and feature preparation for Pricing Model
- Data Formatting and Cleansing for Bundling Model
- Statistical Analysis of Data (Bundling)
- Exploratory Data Analysis (Bundling)

4.Week 7-8:

- Build Initial Model for Product Bundling
- Evaluate Model through different techniques
- Data Formatting and Cleansing for Pricing Model
- Statistical Analysis of Data (Pricing)
- Exploratory Data Analysis (Pricing)

5.Week 9-10:

- Finalize Model for Product Bundling
- Finalize data collection and feature preparation for Pricing Model
- Develop Initial Model for Pricing
- Set and evaluate models through different techniques

6.Week 11:

- Finalization and Fine tuning of both models
- Create a presentation to demonstrate the outcomes

Data Collection:

Identify Relevant Data Sources: Firstly, I identified the specific tables or databases within the bank that contain the required data for my research. I will use transaction details, customer information, product ownership, and account balances tables for data collection.

Legal and Ethical Compliance: I have obtained necessary permissions and ensure that I adhere to legal and ethical standards. I have anonymized or encrypted the data where required. I will then upload a synthetic sample of the dataset to GitHub without any personal information.

Data Collection for Bundling Model: I extracted raw transactional data for each service (merchant category) that I'm interested in. I will perform formatting and initial preprocessing on these datasets to prepare them for the next steps in my analysis.

Link to dataset (anonymized and synthetic sample):

https://github.com/ADA-GWU/guidedresearchproject-Musadiq-Aliyev/tree/main/data

Link to SQL queries (table names replaced by me):

https://github.com/ADA-GWU/guidedresearchproject-Musadiq-Aliyev/tree/main/code

Data Collection for Pricing Model: I will use the raw data to create feature tables for the pricing model. These tables will contain information on customer characteristics such as transactional behavior, personal information, balance status, etc.

Data Cleansing Approaches:

During the data collection phase, I encountered several challenges and issues that required data cleansing. The problems can be grouped as follow:

Missing Data: Since I am using transactional data for the bundling model, there are some customers who do not have transactions on certain services (merchant categories). This could potentially impact the accuracy and reliability of my analysis. To handle this, I will fill the missing values with 0, as it means no transactions.

For the data that I have prepared for the pricing model, I am going to apply techniques such as imputation, where missing values will be estimated based on the available data or replaced with representative values. This approach will ensure that I have complete and meaningful data for accurate analysis in the pricing model.

Inconsistent Formatting: In some cases, I noticed that the data exhibits inconsistencies in formatting. These inconsistencies include variations in date formats and inconsistent data types. To address this issue, I will take steps to standardize the formatting across the dataset. By ensuring consistency and compatibility, I can facilitate accurate analysis and meaningful comparisons.

Outliers: I am aware that outliers can have a significant impact on statistical analysis and modeling results. To identify outliers within my dataset, I will use visualization methods like box plots and statistical measures such as z-scores. Given the large volume of data I have, once outliers are detected, I will proceed to remove them from my dataset.