Case Synthesis

Part1: Establishing a Process Framework

a. Agile coordination framework and life cycle framework

Agile Coordination framework **– Scrum**

Reason:

- It's a well-structured approach to manage work and transparency.
- Well defined roles – Product Owner, Scrum Master, Development team.
- Incremental development and delivery.
- Flexibility and adaptability.
- Coordination with users and working on feedback.
- Sprint Plannings, Daily stand up meeting, Retrospective.

Life Cycle Framework **– CRISP-DM**

Reason:

It is specially designed for data science projects

It also has a well-structured approach –

- Business understanding – Produce project plan to understand the customer requirements and resources required for the project.
    - Identifying leading indicators of customer churn
    - Compare industry average churn to BigNorth Churn to see if the company is an outlier
    - Develop a predictive model with at <85% accuracy for predicting customer churn
    - Create a dashboard to see how churn the company is performing against churn indicators
    - Determine tools and resources that need to be made available to the data science team for this project
- Data understanding – Make sure we have the data flowing that meets the project requirements.
    - Look over the data available and identify if any gaps exist between what we have and what is needed for our models.
    - Make sure the integrity and quality of the data is upto what is expected for this project.
    - Explore the data for redundancy and see if any variables stick out with correlation to churn
- Data preparation – Data collection, Data cleaning and munging.
    - Select the most important variables that were indicated to be relevant to churn
    - Document and format the data so that it is clearly usable by all members of the team
    - See if combining any of the attributes can produce new insight on the data
- Modeling – Determines which algorithm to try and generate test results.
    - Assign models to the teams to compare different architectures
    - Compare the benefits of categorical vs regression
- Evaluation - use the test results and evaluate alongside users' questions.
    - Take successful models and see if any variables are weighted more heavily and if anything is discarded.
    - Decide if the more modeling needs to be done or if the models are ready to be deployed

- Deployment – Deploy the project to the customer and monitor any issues and discuss within the team the challenges faced during the project phase and room for improvement.
    - Plan deployment to the company
    - Review the process for any shortcomings or issues that slowed the project also discuss what went well to ensure it is an aspect that the company keeps in future projects.

Integrating Life Cycle with Agile Framework

- Breaking down the structure of crisp-dm into sprints with daily scrums
- A daily scrum during data understanding could be a horizontal slicing of collecting and cleaning data from many sources
- A daily scrum could build models using the cleaned data which would be vertical slicing.
- Each sprint ensures that the product is a usable increment on the overall.

b. Prior to your first iteration. Some examples are:

Before the first iteration the team will

- Create an item backlog with at least five items, prioritizing tasks such as understanding the business problem, collecting data, and cleaning data.
- Define the Kanban board with columns such as To Do, In Progress, Review, and Done, and set WIP limits to manage the workflow effectively.
- We would organize the 12 employees into sprints to keep the size of the groups below 10 people and assign the roles of scrum master, project owner and development team.
- Decide the tasks that will be assigned to each of the teams.

Goals:

- A detailed business problem statement and a list of measurable project goals.
- Define the tasks to complete
- Conduct stakeholder interviews to gather detailed requirements and understand their expectations
- Document business requirements and success criteria
- Review existing documentation and data related to the business problem to understand the current state and historical context
- Synthesize gathered information into a comprehensive business problem statement.
- Collect information about the people who responded to the NPS surveys.

Priority List:

- Understand the Business Problem
- Data Collection and Understanding
- Data Cleaning and Preparation
- Model Development
- Deployment Planning

Define Kanban Board

Columns:

- Backlog: All the tasks created and ready for pick up in respective sprints
- To Do: List tasks that are planned but not yet started.
- Doing: Tasks that the team is currently working on.
- Testing: Tasks completed but pending review.
- Done: Tasks that are completed and approved.

WIP Limits:

- To Do: Unlimited (all tasks to be done).
- Doing: Limit to 3 tasks per team member to prevent overload.
- Testing: Limit to 5 tasks to ensure timely review and feedback.
- Done: Unlimited (all completed tasks).

Part II: Use Your Process

First Iteration Using the Framework

Goals / Deliverables of the First Item

Goal: Clearly define the business problem and project goals for the Customer Churn case at BigNorth Airlines.

Deliverables:

- A comprehensive business problem statement.
- A list of measurable project goals.
- Documentation of stakeholder requirements and success criteria.

Tasks to Complete the First Item

Create:

- Conduct stakeholder interviews to gather detailed requirements and understand their expectations.
- Document the business requirements and success criteria.

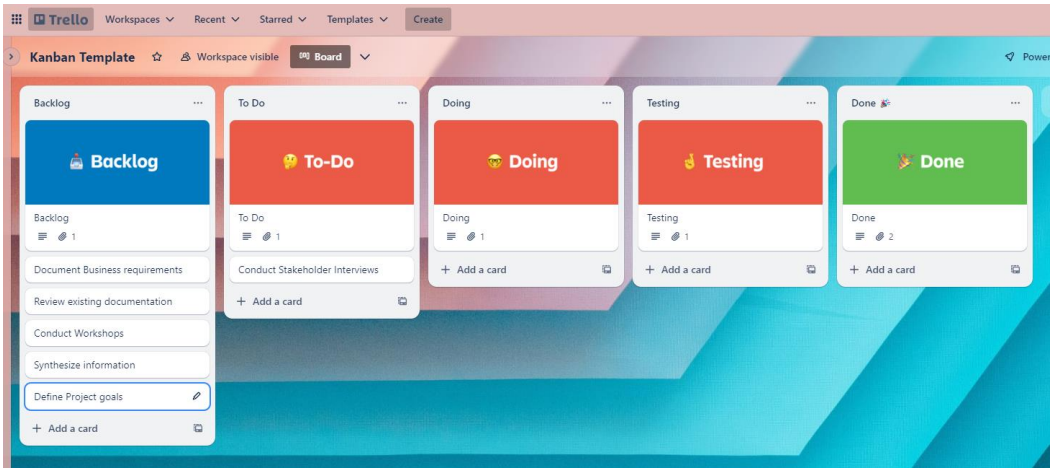
Observe:

- Review existing documentation and data related to the business problem to understand the current state and historical context.
- Synthesize gathered information into a comprehensive business problem statement.

Analyze:

- Analyze the information from stakeholder interviews and documentation to identify key business problems and objectives.
- Validate the business problem statement and project goals with stakeholders.

Kanban Board Visualization:



Communication with the rest of the team

- The team would be assigned to tasks and would have to document findings
- During daily standups, team members will update each other on the progress of these tasks
- The Kanban board will be maintained on a project management tool
- At the end of the sprint, a review meeting will be held to present the completed business problem statement and project goals to the stakeholders for feedback.

Definition of Done:

- Stakeholder interviews will be documented and once all stakeholders have responded to the relevant questions the team will know this step has been accomplished
- Documentation: The business problem statement and project goals are documented in a clear, comprehensive format.
- Review: The documentation is reviewed and approved by key stakeholders, ensuring it meets their requirements and expectations.
- Communication: The final documents are shared with the entire team and uploaded to the project repository for future reference.

Review the Results - first Iteration:

- Look for trends within the interviews to see if any items stand out.
- Conduct a Sprint Review meeting where the team presents the completed business problem statement and project goals to stakeholders.
- Gather feedback from stakeholders on the clarity, completeness, and relevance of the business problem statement and project goals.

Assumption:

- Stakeholder Availability and knowledgeability
- Data is clear, populated, and accessible

Second Iteration:

Goal/Deliverable of the Second Item:

Question:

- What data is relevant and necessary to address business problems and project goals?

Deliverable: A detailed data collection plan and an initial dataset containing relevant data.
Tasks to Complete the Second Item

Create:

- Identify potential data sources required to address the business problem.
- Develop a data collection plan, including a timeline and responsibilities.

Observe:

- Collect sample data from identified sources to assess data availability and quality.
- Review the collected sample data to understand its structure, content, and relevance.

Analyze:

- Evaluate the collected data for completeness, accuracy, and consistency.
- Document the initial dataset and data collection findings, highlighting any data quality issues or gaps.

Questions:

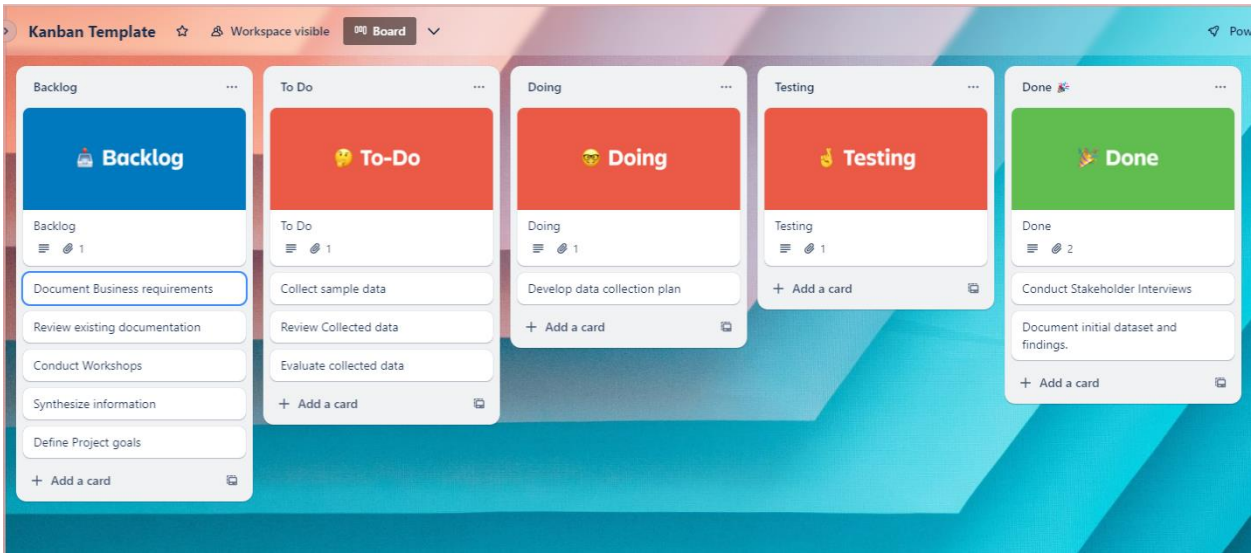
- What are the ethical concerns with the data collected and how it's used?

Deliverable: A report containing what concerns arise from the data collected and how we plan to address bias throughout our analysis.

Create:

- List of variables used in project and assign them a risk factor in relation to the bias they could introduce in the project
- Report on the factors and how the bias could be mitigated or reduced.

Kanban Board Visualization



Communication with the Team

- During daily standups, team members will update each other on the progress of these tasks, discuss any blockers, and plan their activities for the day.
- Real-time communication will be facilitated through a messaging platform like Slack or Microsoft Teams, where team members can ask questions, share updates, and collaborate asynchronously.
- The Kanban board will be maintained on a project management tool, where each task is tracked. Team members can update the status of their tasks and add comments or attachments as needed.
- At the end of the sprint, a review meeting will be held to present the data collection plan and initial dataset to the stakeholders for feedback.

Definition of Done

- **Documentation:** The data collection plan and initial dataset are documented in a clear, comprehensive format.

- **Review:** The documentation and dataset are reviewed and approved by key stakeholders, ensuring they meet the project requirements.
- **Communication:** The final documents and dataset are shared with the entire team and uploaded to the project repository for future reference.
- Reviewing the Results
- **Sprint Review Meeting:** Conduct a Sprint Review meeting where the team presents the data collection plan and initial dataset to stakeholders. This presentation will include a summary of the data sources identified, the data collection process, and the initial findings.
- **Feedback Collection:** Gather feedback from stakeholders on the relevance and quality of the collected data. Discuss any suggestions for additional data sources or improvements in the data collection process.
- **Retrospective:** Conduct a Retrospective meeting with the team to discuss what went well during the sprint, what challenges were faced, and how the process can be improved for future iterations. Document any lessons learned and action items for continuous improvement.

Assumptions

- Relevant data sources are accessible and contain the necessary information to address the business problem.
- Stakeholders are available to provide guidance and feedback on data requirements and collection plans.
- Team members are familiar with the tools and processes used for data collection, evaluation, and documentation.
- The initial data collected is of sufficient quality to begin analysis and identify any major issues early in the process.

By completing these tasks in the second iteration, the team will ensure that they have a solid understanding of the data required for the project and have begun collecting and evaluating that data. This foundational work will support the subsequent phases of data preparation, modeling, and analysis.