

# Career Accelerator

---

## Sprint Planning (LP2)



# In this Sprint Planning

**1** Scrum application in Career Accelerator



**2** Introduction to Project – Classification Project (Customer Churn Prediction)



**3** Assessment Rubric



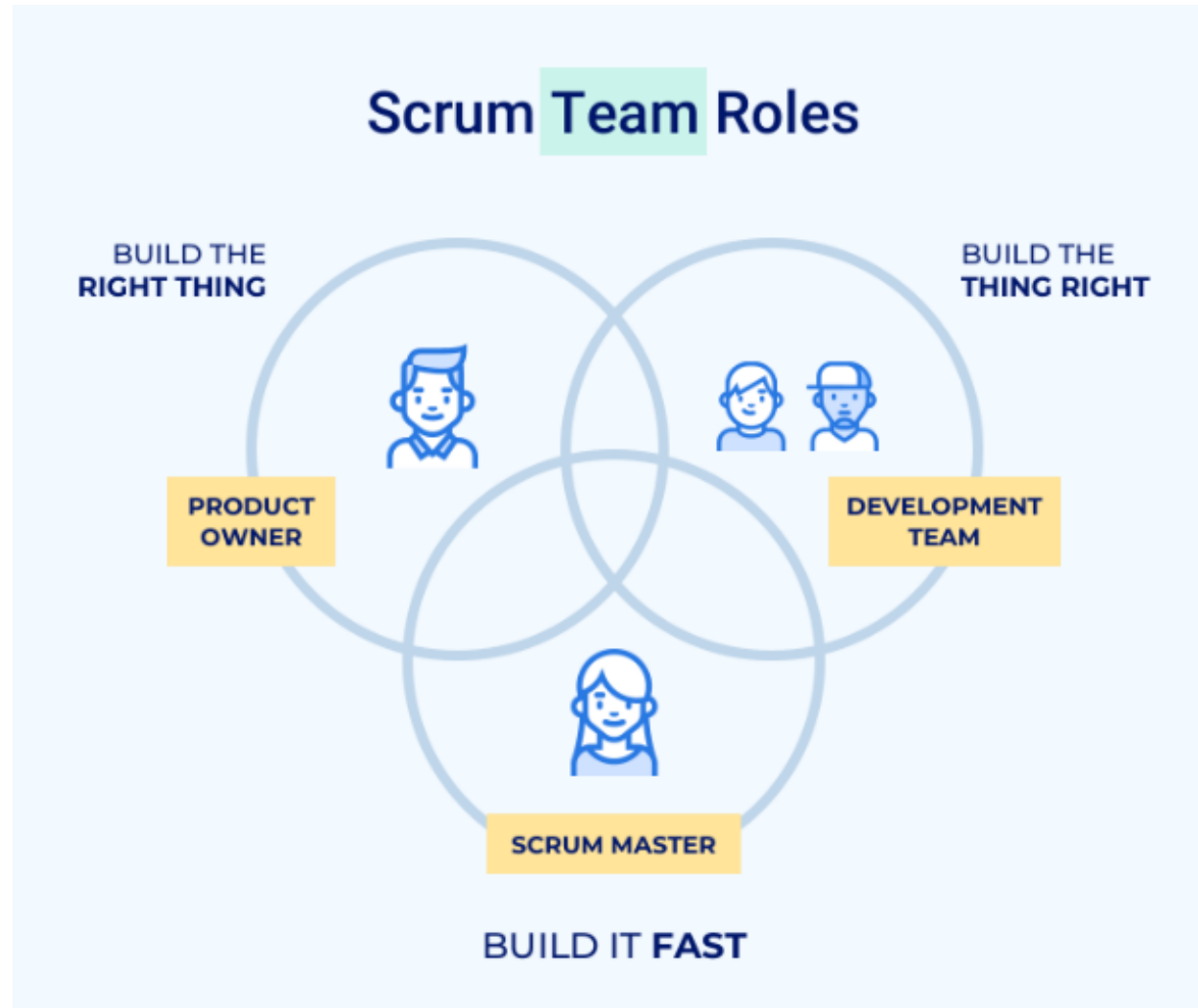
**4** Timeline



**5** How to get help



# The Scrum Approach



# Scrum Artifacts

There are three key Scrum artifacts

1. **Product Backlog:** It contains all the work that needs to be done for the product being developed. **(Team Goal)**
2. **Sprint Backlog:** The sprint backlog is created during the sprint planning meeting from the product backlog as a list of things to be achieved during the current sprint. **(Sprint Goal)**
3. **Increment:** This is the sum of all the product backlog items completed during the sprint and how it adds up to achieve the product goal **(Team goal)**

## Scrum artifacts applied to Career Accelerator

1. **Team Backlog:** 5 projects to make all learners proficient data analysts at the end of the five months.
2. **Sprint Backlog:** Each project is a sprint with tasks to be accomplished towards the sprint goal **(LP2)**.
3. **Increment:** The progression of projects makes learners better analysts with each project.

# Scrum in Career Accelerator

- **What is a Sprint?**
  - Sprints are fixed-length events of one month or less to create consistency during which a development team works on delivering a product (**project**).
- **To foster Scrum during every Sprint**
  - Team members will have a weekly ***Team Sync (Scrum Habits)*** meeting with scrum masters to access the progression of the team.
- **Tool to be used**
  - Kanban Board

# Why is this Sprint Valuable?

1. Learn to work with data from multiple sources for real-world projects
2. Gain relevant skills in data cleaning and exploration
3. Learn how to build classification models that can generalize well
4. Learn to improve and advance your model and test predictions
5. Learn to create robust Machine Learning pipelines

# Objectives

**Sprint Goal:** Build machine learning models that will leverage the dataset and all your data manipulation to predict if a client will **churn** or **not**.

- At the end of the period, you must submit a standalone notebook (ready to be executed by clicking the “run all” button and containing all your technical work) and your article (summarizing your work).
- The main evaluation will be based on the two submitted

 documents.

# Sprint : Customer Churn Prediction

- **Scenario:** You are a data scientist at **Vodafone Corporation**, a large telecommunication company.
- Vodafone wants to find the likelihood of a customer leaving the organization, the key indicators of churn as well as the retention strategies that can be implemented to avert this problem.
- To do this, the business development unit has provided you with data to build a series of machine-learning models to predict customer churn.
- The marketing and sales team have provided you with some data to aid this endeavor.
- Your team uses CRISP-DM Framework for Data Science Projects



# Customer Churn Prediction

Task	Framework Stage
Understanding the problem	Business Understanding
Clean, analyse & visualise the data	Data Understanding
Feature engineering	Data Preparation
Build ML models	Modelling
Evaluate models	Evaluation
Model Improvement	



# Sprint Backlog (Project Requirements)

1. Prepare data for the project (Get data from multiple sources)
2. Develop a hypothesis & ask 5 or more analytical questions
3. Exploratory Data Analysis & Data Cleaning
4. Answer analytical questions with visualizations
5. Deployment of visualizations to Power BI
6. Research on balancing dataset & feature engineering
7. Modeling and Evaluation : Train and evaluate 4 or more models
8. Advanced Model Evaluation
9. Model Improvement
10. Test data Predictions
11. Compile the entire process in an article on medium



# Project Rubrics

- **Documentation:**

**Excellent:** Have documentation on the project i.e., properly document data cleaning, analysis, hypothesis and modelling.

**Good:** Provide a summary on some of the processes

**Fair:** Provide a bullet list of the processes with short sentences

# Project Rubrics

- **Hypothesis & Data Processing:**

**Excellent:** State hypothesis and asked relevant questions. Import all relevant libraries and conduct all checks to make the data ready for future analysis.

**Good:** State hypothesis and implement few steps to check the quality of data.

**Fair:** Don't perform adequate steps in preprocessing the data and also ignore hypothesis

# Project Rubrics

- **Exploratory Data Analysis (EDA):**

**Excellent:** Perform more than 5 bivariate and multivariate analysis coupled with graphs to answer questions and form hypothesis.

**Good:** Perform few bivariate and multivariate analysis coupled with graphs to answer some questions and hypothesis.

**Fair:** Perform only 1 or 2 bivariate & multivariate analyses but omit key visuals like the correlation matrix. Lack of clarity on whether the hypothesis was true or false.

# Project Rubrics

- **Modeling and Evaluation:**

**Excellent:** Train 4 or more models and compare their combination of accuracy, precision, recall and F1 & F2 scores.

**Good:** Train less than 4 models and compare their accuracy, precision, recall and F1 & F2 scores.

**Fair:** Train only 1 model and don't compare any accuracy, precision, recall and F1 & F2 scores.

# Project Rubric

- **Advanced Model Evaluation:**

**Excellent:** Evaluate models through k-Fold cross-validation and explain the rationale for doing so. The results are further visualized on a graph and via Confusion Matrix with detailed explanation of Confusion Matrix visual.

**Good:** Train less than 4 models and compare their accuracy, precision, recall and F1 & F2 scores.

**Fair:** Train only 1 model and don't compare any accuracy, precision, recall and F1 & F2 scores.

# Project Rubric

- **Model Improvement:**

**Excellent:** Perform Hyperparameter tuning and explain the concept into details.

**Good:** Perform Hyperparameter tuning with little explanation on the concept.

**Fair:** Perform Hyperparameter tuning.



# Project Rubric

- **Test data Predictions:**

**Excellent:** Predict on test sets and visualized the results.

**Good:** Predict on test data and only visualized on Confusion Matrix without any explanation of the graph.

**Fair:** Only predict on test data.

# Sprint (4 Weeks)

	MON	TUE	WED	THUR	FRI
Week 1	<b>Sprint Planning</b> (5-7pm GMT)		Team Sync Call (9 – 9:15am GMT)	Engineering Demo (5-7pm GMT)	
Week 2	Team Demo (5-7pm GMT)		Team Sync Call (9 – 9:15am GMT)	Engineering Demo (5-7pm GMT)	
Week 3	Team Demo (5-7pm GMT)		Team Sync Call (9 – 9:15am GMT)	Engineering Demo (5-7pm GMT)	
Week 4	Team Demo (5-7pm GMT)		Team Sync Call (9 – 9:15am GMT)	<b>Sprint Retrospective</b> (5-7pm GMT)	

# Sprint Timeline(4 Weeks)

## Calendar Timeline :

**Monday 10<sup>th</sup> June – Sunday 7<sup>th</sup> July 2024**



# How to get help

These projects are designed to get you to do extensive research on the topic at hand and so we expect you to go out and find answers. However, if you really need to speak/or ask an instructor a question outside of engagement times,

- Please post all your questions in the dedicated channel. i.e. the channel with the name of the current project and the issues' log.
- Additionally, please tag the instructor and give a descriptive assessment of the problem with good screenshots.
- Please utilize the materials curated for you on the Thinkific platform.
- Finally, please fill out the **weekly surveys**, and let us know how we can improve the program for you.