Workplace Overview | Workplace Introduction

Contents

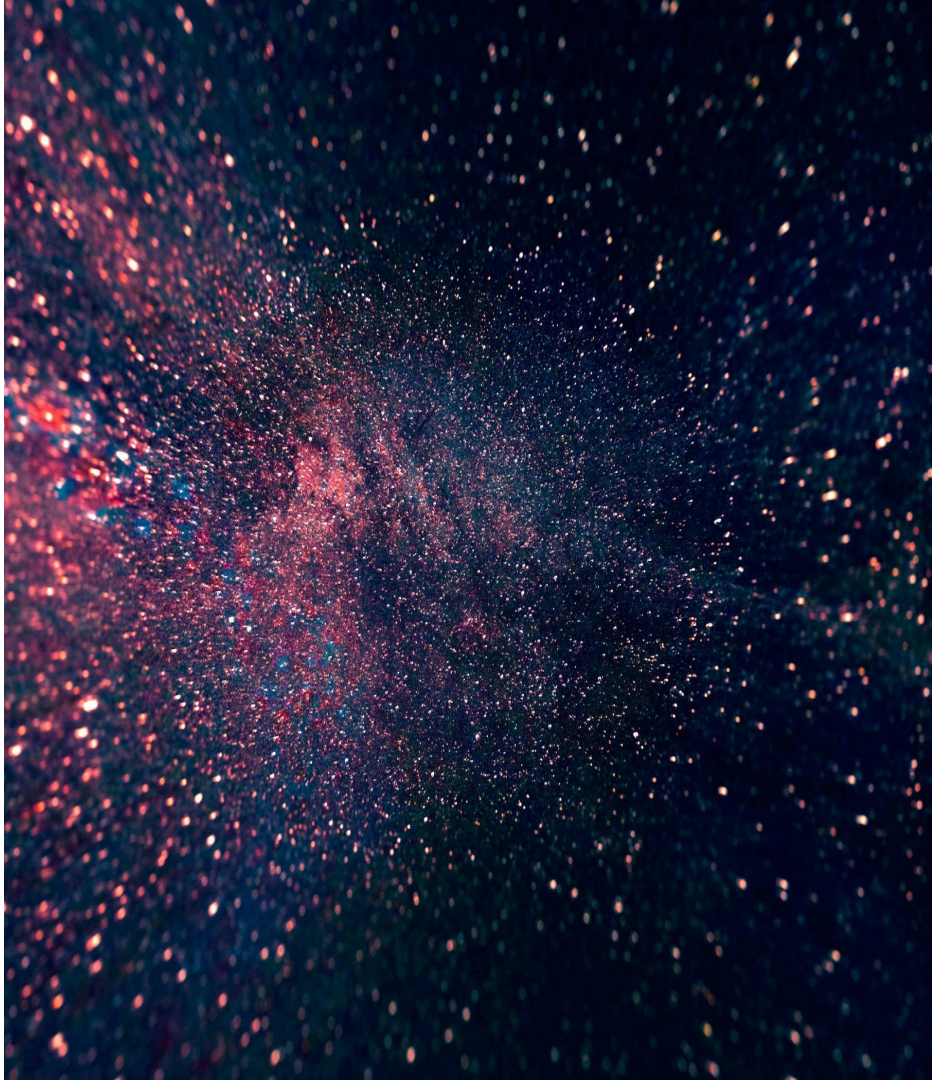
Workplace Introduction

Welcome to the Workplace

Building on the foundations that you have learned throughout the course, this sprint is designed to equip you with the essential skills required in a professional data science workplace environment.

To thrive in a dynamic data science landscape, you will develop problem-solving skills, technical proficiency, effective communication, and adaptability through theoretical knowledge and real-world applications. The topics covered here include data collection and preprocessing, statistical data analysis, data visualisation and reporting.

For this sprint, you will be provided with recorded walk-throughs, interactive notebooks, case studies and informative slides to facilitate your progress and completion of the course. These components will further enhance your learning experience, ensuring you have the necessary support to apply what you have learned in real-world scenarios effectively.



Welcome to the Workplace

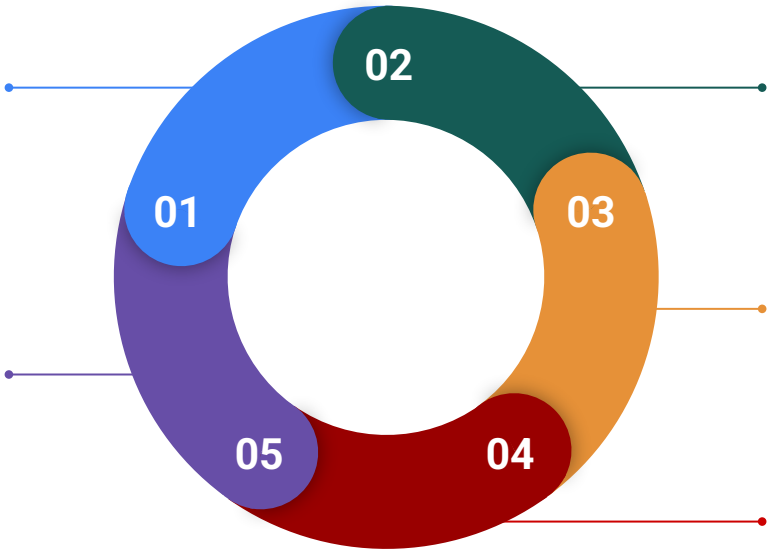
During this **Workplace sprint, our focus will revolve around modules designed to adhere to the data science cycle**. The data science life cycle is a process where we start with understanding a problem and use data to find solutions, including steps like collecting and analysing data to improve business decisions, a key part of business intelligence.

Problem Identification and Business Understanding

Identifying problems and understanding the business context is crucial for developing effective Customer Relationship Management strategies. This aids in pinpointing key issues and aligning data-driven solutions with organisational goals.

Communication of Findings

Communication of Findings involves effectively implementing data-driven solutions and sharing insights with relevant stakeholders to inform decision-making and strategy development.



Data Acquisition, Preparation and Cleaning

Data acquisition, preprocessing, and cleaning involve collecting raw data, preparing it for analysis by correcting errors, and removing irrelevant information to ensure accuracy and reliability.

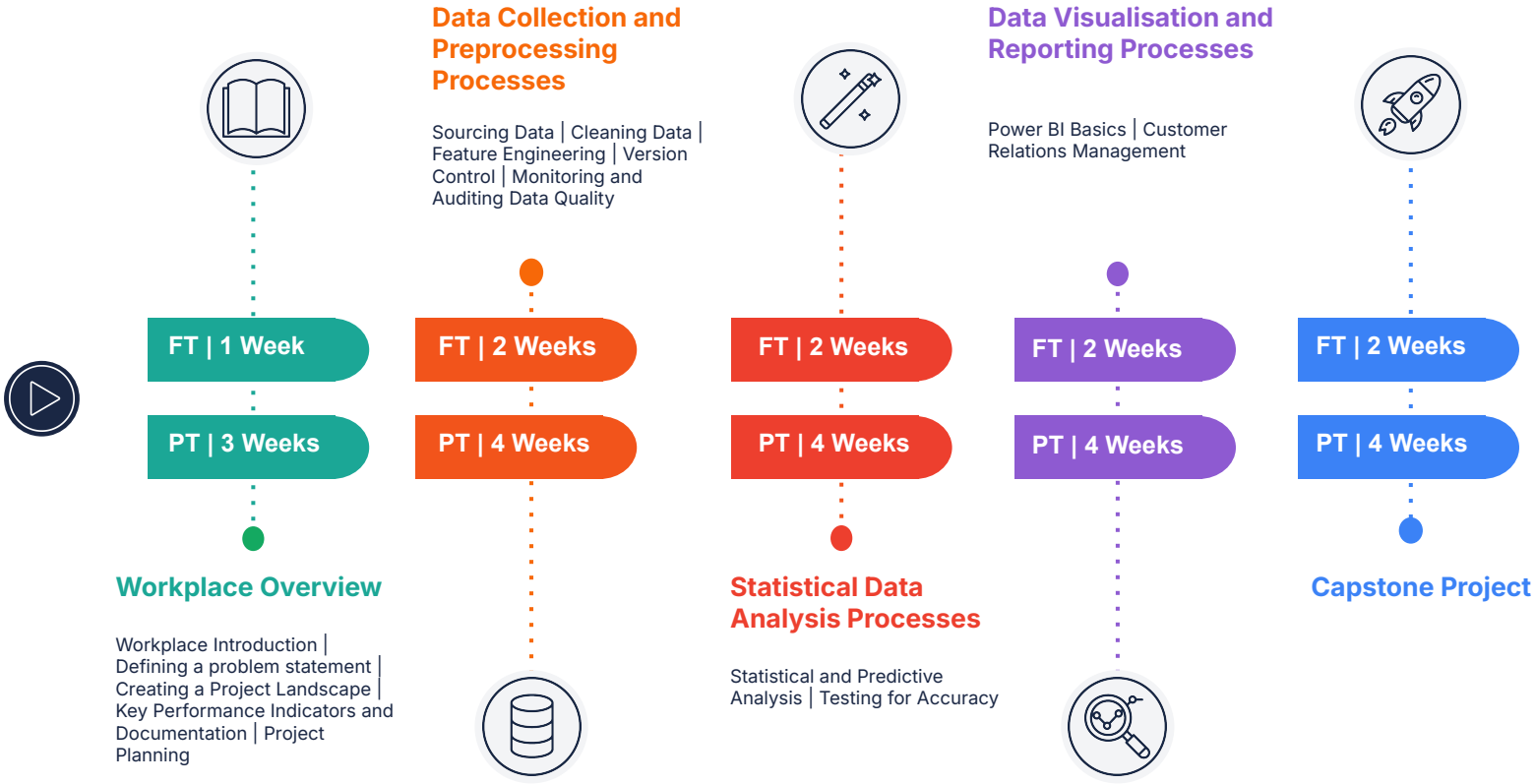
Exploratory Data Analysis

Exploratory Data Analysis is aimed at understanding the main characteristics of the data through visual and statistical methods.

Model Building and Evaluation

Model building and evaluation involves creating predictive algorithms and assessing their accuracy and effectiveness.

Lesson Journey | Workplace Modules



Assessments | Breakdown

During the course of this sprint, we will be focusing on **several critical deliverables.** It is crucial that these deliverables are completed as per the outlined specifications and timelines. Upon completion, all deliverables must be **uploaded to the Athena platform.**

Workplace Overview

- Design project landscape
- Design project board

Data Collection and Preprocessing Processes

- Sourcing data
- Cleaning data
- Feature engineering
- Version control
- Monitoring and auditing data quality

Statistical Data Analysis Processes

- Statistical and predictive analysis
- Testing for accuracy

Data Visualisation and Reporting Processes

- PowerBI dashboard
- Communicate findings
- Customer Relationship Management

Capstone Project

- Project Report

Assessment | Formats

Deliverables are presented in three general formats, each designed to assess our understanding of the content we've previously covered. Make sure to submit these on time and work diligently on each one, as they build upon one another, deepening your comprehension and skill set with each assessment.

Practical Submissions

Practical assessments submitted on Athena could be submitted in the form of slide decks, notebooks, and/ or GitHub repository links. This will allow you to showcase your technical skills and present your work in a professional and accessible format.

Case Studies

Case studies will require you to apply the knowledge that you've learned to hypothetical scenarios. Each case study will be presented in the form of essay questions. Ensure to carefully analyse each scenario and demonstrate your understanding and application of the relevant concepts.

Multiple Choice Questions (MCQ)

MCQs will be a key part of your assessments, designed to test your comprehensive understanding of the course material.

Learning Objectives

Navigating Organisational Data Ecosystems	Learn how to navigate organizational data ecosystems to find and use valuable insights for decision-making.
Implementing Best Practices and Standards in Data Management	Learn how to apply best practices and standards in data management to improve data quality and meet compliance requirements.
Proficient Use of Data Analysis Tools and Platforms	Gain proficiency in using data analysis tools and platforms to efficiently process and interpret data.

Datasets

India Agricultural Dataset :

- https://www.kaggle.com/datasets/vineetkukreti/indian-agriculture-dataset
  Ref: https://doi.org/10.21421/D2/XFB1BZ

Vegetable prices :

- https://www.kaggle.com/datasets/ksamiksha19/vegetable-prices
- Dataset on veggie prices for 10 types over a year.

Global Deforestation :

- **World Forest Area:**
- https://www.kaggle.com/datasets/webdevbadger/world-forest-area
- **More complex set:**
  https://www.kaggle.com/datasets/karnikakapoor/global-forest-data-2001-202

Datasets

Water :

- This research team has extracted water samples from a widely known polluted river located in Buenos Aires, and has analysed them in a chemistry laboratory.
- https://www.kaggle.com/datasets/natanaelferran/river-water-parameters?select=River+water+parameters.csv

Crop Prices :

- Avocado prices over 5 years in USA
- https://www.kaggle.com/datasets/vakhariapujan/avocado-prices-and-sales-volume-2015-2023/code

You may choose your own dataset!

Remember

Choosing a Dataset:

- Ensure the data is clean; consider alternative methods for verifying outliers and imputations.
- Focus on numerical data and try to avoid text-heavy datasets.
- Select a dataset that is both fun and relevant to your interests.
- For a challenge, choose a dataset related to your domain.

Please do not follow the pacing discussed in the video content in the Case Study companion lesson. The training is relevant the dates are not.