

Project Brief

Project title	Worldwide COVID-19 Data Analysis
Module Name	PAI
Course Name	BDSE
Project Start date	Refer Schedule
Project Submission Date	Refer Schedule

Index

1. Purpose of this Project
2. Project Pre-requisites
3. Project Outcomes
4. Project Definition
5. Project Task List
6. Project Evidence
7. Project Guidelines
8. Project Technical Environment
9. Structure of Project Report

1. Purpose of this Project

This Project is used for Summative Assessment of Learner in the Module '**Python for AI**' of the *BDSE*

This Project considers the skills required to Analyse and visualize data using Python programming.

2. Project Pre-requisites

- You should have completed the following activities before.
 - Viewed and understood all the e-content related to the module.
 - Completed all the MCQ tests related to the module.
 - Completed all the Assignment / Lab Exercises of the module.
- You should have access to the Project Brief, Project Report template and should understand how to use the templates.
- You have access to the Project Technical Environment
- You should understand the number of milestones and what are the milestones to be presented for each of the Tutoring Session
- To implement the project, you will need a MS Windows Computer/Mac on which Anaconda, Python, Azure data Studio and Visual Studio code is installed, or you can use google colab environment.

3. Project Outcomes & Deliverables

- You should perform all the tasks in the project task list and prepare the following during the project:
- Implement the project on Python and prepare .py files or. ipynb files
- Prepare a project report as per the pre-defined template.

4. Project Definition

The global outbreak of the COVID-19 pandemic has underscored the critical need for accurate and timely health data analysis. ABC Health Analytics Company recognizes the significance of leveraging data to understand and respond to the evolving challenges posed by the pandemic. In this context, the company is initiating a project aimed at comprehensive data analysis and visualization of COVID-19 datasets using Python. The goal of this capstone project is to perform a comprehensive analysis of worldwide COVID-19 data. We will leverage data visualization techniques using Python libraries such as Pandas, Matplotlib, and seaborn. The dataset used for this analysis is sourced from "Our World in Data" and contains information on various COVID-19 metrics across different countries.

Here you'll focus on Pandas, Matplotlib, and Seaborn aspects to understand, analyse, and visualise the data.

5. Project Task List

For the above case study, perform the following project activities:

Activity 1: Data Loading and Exploration

Initializing the analysis by loading and exploring the dataset, ensuring data integrity and format consistency for accurate downstream analysis.

1. Load the COVID-19 DataFrame using pandas from the provided dataset files.
2. Display the first 5 and last 5 rows of the DataFrame.
3. Check for missing values in the dataset and decide on a strategy to handle them.
4. Remove the columns with more than 90% of missing values.
5. Convert the 'date' column to the datetime data type.

Activity 2: Data Cleaning and Feature Engineering

Enhancing data relevance and structure, enabling focused analysis on key COVID-19 indicators and demographic factors.

1. Impute missing values in the dataset columns
2. Remove duplicate rows from the DataFrame.
3. Create new features if needed (e.g., extract year and month from the 'date' column).
4. Explore unique countries in the dataset and count the total number of countries.

Activity 3: Worldwide COVID-19 Overview

Providing a global and temporal perspective on COVID-19, identifying patterns and correlations crucial for strategic decision-making.

1. Visualize the WHO Regions with total COVID-19 cases and total deaths by using bar plots.
2. Explore the world wide monthly trend of COVID-19 cases using a line plot.
3. Investigate the correlation between total cases and total deaths using a heatmap..
4. Analyze how total cases have evolved over time for a specific location (e.g., India).

Activity 4: Regional Analysis

Unveiling regional nuances in COVID-19 dynamics and identifying patterns in case distribution and fatality across time and continents.

1. Create a grouped bar chart to visualize new cases by continent and month.
2. Visualize the distribution of total COVID-19 cases by year using a box plot.
3. Compare total deaths across different continents using a bar plot.
4. Analyze the total cases on a month-by-month basis using a bar plot.

Activity 5: Time Series Analysis

Capturing the daily dynamics of COVID-19, assessing the impact of vaccination, and monitoring testing metrics for a comprehensive global overview.

1. Explore the daily trend of new cases and new deaths globally using line plots.
2. Calculate and visualise the daily average of new cases and deaths globally.
3. Explore the trend of vaccination coverage over time globally.
4. Analyze the total tests and positive rate over time globally.

Activity 6: In-Depth Country Analysis

Conducting an in-depth examination of specific countries, understanding the correlation between cases and deaths globally, and exploring continental variations in case distribution.

1. Visualize a specific country's total COVID-19 cases and deaths over time.
2. Get user input for the country and metric (total_cases or total_deaths) and plot a line chart for the user-selected values.
3. Analyze the distribution of total cases across different continents using a box plot.
4. Visualize the Year wise monthly trend of new COVID-19 cases for user's selected country.

Activity 7: Additional Insights

Extracting additional insights, examining the influence of external factors, and evaluating regional disparities for a holistic understanding of the COVID-19 landscape.

1. Visualize the fatality rate (total deaths / total cases) over time globally.
2. Explore the positivity rate (total_cases/ total tests) versus total tests conducted to analyse testing effectiveness using the x-axis as the logarithmic scale for better visualisation.
3. Analyze the fatality rate and its relationship with smoking (Use male_smokers and female_smokers columns).
4. Create a heatmap to analyse the relationship between hospital beds per thousand and fatality rate.

6. Project Evidence

The Learner must submit the following evidence.

A Project Report which comprises of the screen shots of each activity to show that the Tasks of each activity has been executed correctly.

7. Project Guidelines

You should follow the below guidelines while implementing the Project:

- Implement the project in the technical environment specified in the Project brief.
- Follow the format specified for Project Report
- The project report should be submitted at least 1 days before the date of Summative Assessment date.
- Present the Milestones in every Tutoring Session and seek the Tutor's feedback and review. Incorporate the feedback in your project.
- Attach all project evidence for each milestone as part of your Project report.

8. Project Technical Environment

The Learner should perform the project using Colab/Jupyter notebook, Python, and the specified dataset.

9. Structure of Project Report

- Index of the Project Report (1 Page)
- Technical environment
- Problem statement
- Activity 1
- Activity 1 screenshot
- Activity 2
- Activity 2 screenshot
- Activity 3
- Activity 3 screenshot
- Activity 4
- Activity 4 screenshot
- Activity 5
- Activity 5 screenshot
- Activity 6
- Activity 6 screenshot
- Activity 7
- Activity 7 screenshot
- Refer the project Template for better understanding.