How-To Guides for Project Plan

Group3\_Project1

La Shawn Sykes

**Day 1 Task: Data Collection**

Research and identify relevant datasets for the project.

Utilize online repositories, academic journals, and data platforms such as Kaggle to search for datasets related to the project's objectives.

Consider factors such as data quality, relevance to the project goals, and availability of necessary variables.

**Day 2 Task: Data Integration and Cleaning**

Combine and integrate the acquired datasets into a single, unified dataset for analysis.

Use tools like Pandas in Python to merge datasets based on common columns or keys.

Ensure the integrated dataset maintains data integrity and consistency.

**Day 3 Task: Exploratory Data Analysis (EDA)**

Perform in-depth analysis of the preprocessed dataset to identify patterns, trends, and relationships.

Utilize descriptive statistics and visualizations to gain insights into the dataset's characteristics.

Consider using tools like Matplotlib and Seaborn to create visual representations of the data.

**Day 4 Task: Final Data Analysis**

Perform a final review and validation of the analysis results.

Cross-verify key findings and insights to ensure accuracy and reliability.

Document the final analysis outcomes for presentation and reporting.

Aaron Cranor

**Day 1 Task: Environment Setup**

Set up the programming environment with required libraries (e.g., Pandas, NumPy, Matplotlib, Seaborn).

Install and configure Python and relevant libraries using tools like Anaconda or pip.

Verify that all necessary packages and dependencies are properly installed.

**Day 2 Task: Data Cleaning**

Handle missing or null values in the dataset and remove duplicate entries and outliers.

Use Pandas functions to identify and handle missing values, such as fillna() or dropna().

Employ statistical techniques and visualization tools to identify and address outliers.

**Day 3 Task: Advanced Data Analysis**

Apply advanced statistical and analytical methods to uncover deeper insights within the dataset.

Utilize techniques such as regression analysis, cluster analysis, or time series analysis as applicable.

Explore relationships between variables using correlation analysis and regression models.

**Day 4 Task: Presentation Preparation**

Compile the analysis results and visualizations into a cohesive presentation format.

Organize key findings and insights into a structured presentation outline.

Create visually appealing and informative slides using presentation software such as PowerPoint or Google Slides.

Lewis Hill

**Day 1 Task: Initial Data Exploration**

Perform a preliminary review of the acquired datasets to understand their structure and contents.

Use Pandas functions to load and examine the dataset's structure, such as head(), info(), and describe().

Identify potential areas for further analysis based on the initial dataset exploration.

**Day 2 Task: Feature Engineering**

Derive and create new features or variables from the existing dataset to enhance analysis.

Engineer new features based on domain knowledge and data understanding.

Consider feature scaling and transformation to ensure uniformity and relevance.

**Day 3 Task: Visual Representation**

Create visualizations such as scatter plots, bar charts, and heatmaps to represent key findings.

Use Seaborn and Matplotlib to generate visual representations of the data.

Select appropriate visualization types based on the nature of the data and the insights to be conveyed.

**Day 4 Task: Documentation**

Prepare a detailed project report with the analysis methodology and results.

Document the data preprocessing and analysis workflow for future reference.

Ensure the project report includes clear explanations and interpretations of the analysis outcomes.