

Welcome to **INTERNSHIP STUDIO**

Final Project

Sales Data Analysis and Reporting for a Retail Chain

Project Plan

The aim of this project is to use Python, SQL, and Excel to analyze sales data and generate meaningful reports for a retail chain.

Data Source

Find the data here: [Retail Data](#). This dataset includes the following fields:

- **TransactionID**: A unique identifier for each transaction
- **TransactionTime**: The time the transaction took place
- **ItemCode**: The code of the item purchased
- **ItemDescription**: A description of the item purchased
- **NumberOfItemsPurchased**: The number of items purchased in the transaction
- **CostPerItem**: The cost per item
- **Country**: The country where the transaction took place

Phase 1: Data Collection and Database Setup

- 1.Data Collection:** Download the data from Kaggle as a csv file and place it on the proper path
- 2.Database Setup:** Set up a SQL database to hold the data. Design the database schema, and create the necessary tables using SQL DDL commands.

Phase 2: Data Cleaning and Preparation

1. **Data Cleaning:** Use SQL queries and Python (pandas) to clean the data. Look for and handle missing or inconsistent data, outliers, etc.
2. **Data Preparation:** Prepare the data for analysis. This may involve creating additional calculated fields, such as total sales value, month/year fields for time-based analysis, etc. Again, this can be done using a combination of SQL and Python.

Phase 3: Data Analysis

1. **Data Exploration:** Use SQL queries and Python (pandas, matplotlib, seaborn, etc) to explore the data and identify trends and patterns.
2. **Advanced Analysis:** Perform more complex analysis as needed. For example, time series analysis for sales trends, cohort analysis for customer behavior, etc. Python's advanced data analysis libraries can be very useful here.

Phase 4: Reporting

1. **Report Preparation:** Prepare reports summarizing the findings. These can include:
 - **Tabular Reports:** Create summary tables showing sales by product, store, month, etc. These can be generated using SQL queries and Python, and then formatted and presented in Excel.
 - **Visual Reports:** Create visual reports (charts, graphs, etc) showing trends and patterns. These can be created using Python's data visualization libraries, and then incorporated into Excel for presentation.
 - **Automated Reports:** If needed, set up automated reports that run regularly and update Excel dashboards, using Python for automation.
2. **Report Presentation:** Present the reports to in either your Excel dashboard or prepare a PPT to present the findings.

ALL THE BEST