# PlatinumRx Assignment Reference Document

## 1. Objective

The goal of this project is to assess and demonstrate core data analysis skills. You will be building a small portfolio of solutions that covers:

- **Database Management (SQL):** Creating schemas and querying data for Hotel and Clinic management systems.
- **Data Manipulation (Spreadsheets):** Using functions to look up data and perform time-based analysis.
- **Programming Logic (Python):** Writing scripts for data conversion and string manipulation.

## 2. Technical Requirements

To complete this assignment effectively, you will need the following tools:

- **For SQL Tasks:**
  - **Software:** A relational database management system (RDBMS) like **MySQL Workbench**, **PostgreSQL**, or an online SQL compiler (e.g., SQLiteOnline, DB Fiddle).
  - **Knowledge:** Understanding of `CREATE TABLE`, `INSERT`, `JOIN`, `GROUP BY`, and Aggregations.
- **For Spreadsheet Tasks:**
  - **Software: Microsoft Excel** or **Google Sheets**.
  - **Knowledge:** VLOOKUP/XLOOKUP, Date/Time functions, Pivot Tables or COUNTIFS.
- **For Python Tasks:**
  - **Software:** Python 3.x installed locally with an IDE (like **VS Code**, **PyCharm**) or an online environment like **Jupyter Notebook** or **Google Colab**.
  - **Knowledge:** Basic syntax, variables, loops (`for` / `while`), and arithmetic operators.

## 3. Project Structure

Organize your files cleanly to keep track of your work. We recommend the following folder structure:

```
Data_Analyst_Assignment/
│
├── SQL/
│   ├── 01_Hotel_Schema_Setup.sql    # Table creation and data insertion for Hotel
│   ├── 02_Hotel_Queries.sql         # Solutions for Part A (Questions 1-5)
│   ├── 03_Clinic_Schema_Setup.sql   # Table creation and data insertion for Clinic
│   └── 04_Clinic_Queries.sql        # Solutions for Part B (Questions 1-5)
│
├── Spreadsheets/
```

```
│   └── Ticket_Analysis.xlsx        # The workbook containing data and analysis
│
├── Python/
│   ├── 01_Time_Converter.py        # Script for minutes conversion
│   └── 02_Remove_Duplicates.py     # Script for string manipulation
│
└── README.md                       # (Optional) Brief notes on your approach
```

# 4. Step-by-Step Instructions

## Phase 1: SQL Proficiency

This section involves two scenarios: a Hotel Management System and a Clinic Management System.

### Step 1: Setup the Database

Before querying, you need to create the "skeleton" of the database.

- **Action:** Write `CREATE TABLE` statements for every table listed in the assignment (e.g., `users`, `bookings`, `items` for the Hotel system).
- **Action:** Write `INSERT INTO` statements to add the sample data provided in the PDF so you can test your queries.

### Step 2: Hotel System Analysis (Part A)

You need to write queries for 5 specific questions.

- **Logic Guide:**
  - *Q1 (Last booked room):* You need to find the latest date. Think about `MAX(date)` or sorting by date and limiting the result.
  - *Q2 (Billing in Nov 2021):* This requires joining `bookings`, `booking_commercials`, and `items`. You will need to calculate `quantity * rate` to get the amount.
  - *Q3 (Bills > 1000):* Use the `HAVING` clause to filter after you have summed up the bill amounts.
  - *Q4 (Most/Least ordered):* You need to group by Month and Item. This might require a "Window Function" (like `RANK` or `ROW_NUMBER`) to find the top and bottom items per month.
  - *Q5 (2nd Highest Bill):* Similar to Q4, use a ranking function to find the bill in the 2nd position.

### Step 3: Clinic System Analysis (Part B)

- **Logic Guide:**
  - *Q1 (Revenue by Channel):* A simple `GROUP BY sales_channel` and `SUM(amount)`.
  - *Q3 (Profit/Loss):* Profit is `Revenue - Expenses`. You will likely need to aggregate Revenue (from `clinic_sales`) and Expenses (from `expenses`) separately by month, then join those results together to do the math.

## Phase 2: Spreadsheet Proficiency

You are given `ticket` and `feedbacks` data.

### Step 1: Data Preparation

- **Action:** Create an Excel file. Create two sheets named "ticket" and "feedbacks". Copy the sample data into them.

### Step 2: Populating Data (Question 1)

- **Goal:** Bring the `created_at` date from the `ticket` sheet into the `feedbacks` sheet.
- **Action:** Use the `cms_id` column as the common link (key) between the two sheets.
- **Tool:** Use `VLOOKUP` or `INDEX-MATCH`.

### Step 3: Time Analysis (Question 2)

- **Goal:** Count tickets created and closed on the **same day** and **same hour**.
- **Action:**
  1. Create a helper column "Same Day?" → Check if `Int(Created_Date) == Int(Closed_Date)`.
  2. Create a helper column "Same Hour?" → Check if the hour part of the timestamp matches.
  3. Use `COUNTIF` or a Pivot Table to count the "True" values per outlet.

## Phase 3: Python Proficiency

### Step 1: Time Conversion

- **Goal:** Convert integer minutes (e.g., 130) into "X hrs Y minutes".
- **Logic:**
  - Hours = Total Minutes divided by 60 (integer division).
  - Remaining Minutes = Total Minutes modulo (%) 60.

### Step 2: Remove Duplicates

- **Goal:** Take a string and remove duplicate characters using a loop.
- **Logic:**
  - Create an empty string variable (e.g., `result = ""`).
  - Loop through every character in the input string.
  - **Check:** If the character is *not* already in `result`, add it. If it is, skip it.

# 5. Example Code (Snippets)

## Excel: VLOOKUP Syntax

Used to pull the Ticket Created Date into the Feedbacks sheet.

```
=VLOOKUP(
    lookup_value,      -- The unique ID (cms_id) in the current row
    table_array,       -- The range in the 'ticket' sheet (e.g., 'ticket'!E:B)
    col_index_num,     -- The column number containing 'created_at'
```

```
    [range_lookup]     -- FALSE (for exact match)
)
```

## 6. Submission

Please ensure that your completed work is properly organized and accessible for review.

### 📁 Required Submission Links

- **GitHub Repository:**

  Provide the link to your GitHub repository containing the following folders:

  - `SQL/` — All `.sql` scripts for both Hotel and Clinic systems.
  - `Spreadsheets/` — Your Excel or Google Sheet file (`Ticket_Analysis.xlsx`).
  - `Python/` — The Python scripts (`01_Time_Converter.py`, `02_Remove_Duplicates.py`).
  - *(Optional)* `README.md` summarizing your approach and assumptions.

- **Spreadsheet Link:**

  Share the Google Sheets link (or upload the `.xlsx` file) used for the spreadsheet tasks in Phase 2.

  Ensure that the link is accessible with “View” permissions for reviewers.

### 🧾 Final Checklist

Before submitting, confirm that:

1. All SQL scripts execute without errors and return expected results.
2. Spreadsheet formulas are working correctly (especially lookups and time comparisons).
3. Python scripts run successfully and handle edge cases.
4. Links provided are publicly accessible or shared appropriately.

### 📩 Submission Format Example

```
**GitHub Link:** https://github.com/yourusername/PlatinumRx_Assignment
**Spreadsheet Link:** https://docs.google.com/spreadsheets/d/yourspreadsheetlink
```