# **IPL Data Analysis Project**

**A complete end-to-end Data Analytics project using Excel, SQL (PostgreSQL), Python (Google Colab), and Power BI.**

**Project Overview**

**Analyze IPL match and ball-by-ball data to uncover insights about teams, players, and match conditions.**

## **Project Structure**

**ipl-data-analysis/  
|  
|── Excel/  
| ├─ cleaned\_matches.csv  
| └─ cleaned\_deliveries.csv  
|  
|── SQL/  
| ├─ matches\_final.csv  
| ├─ deliveries\_final.csv  
| └─ queries.sql  
|  
|── Python/  
| ├─ ipl\_data\_analysis.ipynb  
| └─ ipl\_data\_analysis.py  
|  
|── PowerBI/  
| └─ IPL\_Data\_Analysis\_Report.pbix  
|  
└─ README.md**

## **Tools Used**

* **Excel – Data Cleaning (null values, formatting, pivot tables)**
* **PostgreSQL (psql shell) – Data Storage and Queries**
* **Python (Google Colab) – Exploratory Data Analysis and Visualization**
* **Power BI – Dashboard Creation**

## **Step 1: Installation & Setup**

### **1. PostgreSQL Installation**

* **Download from: <https://www.postgresql.org/download/>**
* **Install with default configurations.**
* **Ensure psql shell and pgAdmin are installed.**

### **2. Power BI Installation**

* **From Microsoft Store: Search “Power BI Desktop” > Install**
* **Or: <https://powerbi.microsoft.com/desktop/>**

### **3. Google Colab (No Installation Required)**

* **Visit: <https://colab.research.google.com>**
* **Use your Google account to run Python notebooks in the cloud**

## **Step 2: Data Preparation (Excel)**

* **Open raw match/delivery CSVs**
* **Clean columns, remove blanks, fix encodings**
* **Save as cleaned\_matches.csv and cleaned\_deliveries.csv**

## **Step 3: SQL Setup & Load**

### **Create Database and Tables**

**CREATE DATABASE ipl\_analysis;  
\c ipl\_analysis  
  
-- Matches Table  
CREATE TABLE matches (  
 id INT PRIMARY KEY,  
 season INT,  
 city TEXT,  
 date DATE,  
 team1 TEXT,  
 team2 TEXT,  
 toss\_winner TEXT,  
 toss\_decision TEXT,  
 result TEXT,  
 dl\_applied INT,  
 winner TEXT,  
 win\_by\_runs INT,  
 win\_by\_wickets INT,  
 player\_of\_match TEXT,  
 venue TEXT,  
 umpire1 TEXT,  
 umpire2 TEXT  
);  
  
-- Deliveries Table  
CREATE TABLE deliveries (  
 match\_id INT,  
 inning INT,  
 batting\_team TEXT,  
 bowling\_team TEXT,  
 over INT,  
 ball INT,  
 batsman TEXT,  
 non\_striker TEXT,  
 bowler TEXT,  
 is\_super\_over INT,  
 wide\_runs INT,  
 bye\_runs INT,  
 legbye\_runs INT,  
 noball\_runs INT,  
 penalty\_runs INT,  
 batsman\_runs INT,  
 extra\_runs INT,  
 total\_runs INT,  
 player\_dismissed TEXT,  
 dismissal\_kind TEXT,  
 fielder TEXT  
);**

### **Load CSVs into SQL**

**\COPY matches FROM 'C:/pgimport/cleaned\_matches.csv' DELIMITER ',' CSV HEADER;  
\COPY deliveries FROM 'C:/pgimport/cleaned\_deliveries.csv' DELIMITER ',' CSV HEADER;**

## **Step 4: Python Analysis (Google Colab)**

* **File: Python/ipl\_data\_analysis.ipynb**
* **Contains:**
  + **EDA using pandas, matplotlib, seaborn**
  + **Top scorers, most wickets, match trends, pie charts**

**Sample Code:**

**import pandas as pd  
import matplotlib.pyplot as plt  
  
matches = pd.read\_csv('/content/cleaned\_matches.csv')  
deliveries = pd.read\_csv('/content/cleaned\_deliveries.csv')  
  
# Top 10 players with most Player of Match awards  
matches['player\_of\_match'].value\_counts().head(10).plot(kind='bar')  
plt.title("Top 10 Man of the Match Winners")  
plt.ylabel("Count")  
plt.show()**

## **Step 5: Connect SQL to Power BI**

### **1. Open Power BI Desktop**

### **2. Click Home > Get Data > PostgreSQL**

### **3. Provide:**

* **Server: localhost**
* **Database: ipl\_analysis**
* **Authentication: postgres user**

**Important: Choose Import mode**

### **4. If Encryption Error Appears:**

* **Click Edit > Disable encryption > Retry**

### **5. Load Tables**

* **Select matches and deliveries**
* **Load the tables into Power BI**

## **Step 6: Build Visuals in Power BI**

| **Chart Type** | **What to Show** |
| --- | --- |
| **Bar Chart** | **Top 10 Batsmen with most runs** |
| **Line Chart** | **Runs scored over seasons** |
| **Pie Chart** | **Toss decisions (bat vs field)** |
| **Donut Chart** | **Number of matches per venue** |
| **Table** | **Match summary table (add filters on team/season)** |

### **Add Titles & Axis**

* **Click the visual**
* **Go to Visualizations > Format (paint roller)**
* **Turn ON Title, X-Axis, Y-Axis**

## **Step 7: Export CSVs from PostgreSQL**

**\COPY (SELECT \* FROM matches) TO 'C:/pgimport/matches\_final.csv' DELIMITER ',' CSV HEADER;  
\COPY (SELECT \* FROM deliveries) TO 'C:/pgimport/deliveries\_final.csv' DELIMITER ',' CSV HEADER;**

## **Credits**

* **Dataset from Kaggle**
* **Visuals by Power BI**
* **Analysis by Google Colab**