

# New Yor City Government Job Postings Public Dataset Analysis in MySQL and Power BI

## Project Overview

The project is structured in four parts and aims to conduct an exploratory analysis of New York City government job postings. It leverages advanced data cleaning and transformation techniques in MySQL and presents insights through a Power BI dashboard with interactive visualizations.

---

## Source

The database was accessed on the DATA.GOV government website, link:  
<https://catalog.data.gov/dataset/nyc-jobs>

Accessed on 5/31/2025.

---

## 1. Preparation of the CSV File for MySQL Import

- **Columns used:**  
B (Agency), C (Posting Type), D (# Of Positions), F (Civil Service Title), J (Job Category), K (Full-Time/Part-Time Indicator), L (Career Level), M (Salary Range From), N (Salary Range To), O (Salary Frequency), P (Work Location), AA (Posting Date)
  - **Columns removed:**  
A (Job ID), E (Business Title), G (Title Classification), H (Title Code No), I (Level), Q–AD (Additional job description and contact details)
  - **Columns renamed:**  
Posting\_Type, Number\_Of\_Positions, Civil\_Service\_Title, Job\_Category, Full-Time\_Part-Time\_Indicator, Career\_Level, Salary\_Range\_From, Salary\_Range\_To, Salary\_Frequency, Work\_Location, Posting\_Date
  - **File renamed to:**  
"Jobs\_NYC\_Postings\_SQL\_Ready"
- 

## 2. Data Cleaning and Transformation in MySQL

- **Remove duplicates:** Using GROUP BY, HAVING, and DELETE statements
- **Handle null/missing values:** WHERE IS NULL, UPDATE, SET

- **Normalize values in Full-Time/Part-Time Indicator:** Change 'F' to 'Full-Time' and 'P' to 'Part-Time'
  - **Standardize salary frequency and remove the column:** CASE WHEN, ALTER TABLE DROP COLUMN
  - **Create salary brackets:** Use 11 defined salary ranges via CASE statements
  - **Group job categories into 14 main groups:** Using LIKE, UPDATE, SELECT DISTINCT
- 

### 3. Data Analysis in MySQL

- **Overall metrics:**
    - Total job postings: COUNT(\*)
    - Total number of positions: SUM(Number\_Of\_Positions)
    - Time range covered in dataset: MIN(Posting\_Date) / MAX(Posting\_Date)
  - **Time-series analysis:**
    - Monthly trend of job postings and positions hired: DATE\_FORMAT(), GROUP BY, ORDER BY
  - **Categorical distributions:**
    - Posting type, employment type, location
    - Top 10 agencies and civil service titles
    - Job categories, salary brackets, career level
  - **Percentages of distributions** calculated using nested SELECT COUNT()/SUM() queries
- 

### 4. Power BI Dashboard Development

- **KPI Cards:**
  - Total job postings
  - Total positions hired
  - First posting date

- **Time-based visualizations:**
  - Line charts for monthly trends in postings and positions hired
- **Distributions:**
  - Pie charts for posting types, employment types, and career level
  - Maps for job posting locations
  - Bar charts for top agencies and titles
  - Tables showing breakdowns by job category and salary bracket with counts and % of total

---

## 5. Project Files

- **Jobs\_NYC\_Postings\_Original.csv** - The original raw dataset of NYC government job postings (CSV)
- **Jobs\_NYC\_Postings\_SQL\_Ready.csv** - Formatted CSV file prepared for MySQL import
- **Data\_Cleaning\_and\_Transformation.sql** - MySQL script for data cleaning, normalization, and transformation
- **Jobs\_NYC\_Postings\_Cleaned.csv** - Exported dataset after cleaning and processing in MySQL
- **Data\_Analysis\_and\_Visualization.sql** - MySQL script containing analysis queries used to generate insights and prepare data for visualization
- **Data\_Visualisation\_Queries** - Folder with Excel files exported from MySQL query results, used as source data for Power BI
- **Jobs\_NYC\_Final\_PoweBI\_Report.pbix** - Power BI project file containing the interactive dashboard and visualizations
- **Jobs\_NYC\_Final\_PoweBI\_Report.pdf** - PDF export of the Power BI dashboard for offline viewing
- **README.pdf** - PDF version of the project README with documentation and instructions