Project 1 Report

Introduction

The dataset we picked from data.gov contains the NYC Job Postings that are available to see on the official site for the City of New York. The data includes listings that are only available to city employees and listings that are open to the public.

Data Inspection

This 37 MB dataset of NYC jobs has 5790 rows of data and 30 columns. Most of the columns are nominal data but there are also columns for quantitative data and date data. 25 of the columns contain objects. 3 of them contain float data. 2 of them contain integers. Most of these listings have only 1 position available. Most of the columns that have nominal data are organizational (agency, role business title, job category, career level). The amount of internal and external listing types are almost completely evenly split. There's ordinal data in the Level column. The salary columns are using floats for the figures.

Most of the columns were self-explanatory. The Job ID column was clearly the primary key for each listing. There are 2 columns for the titles of each listing: Business Title and Civil Service Title. The Title Classification column separates listings as Competitive or Non-Competitive. Competitive means that there's a civil service exam that must be taken and passed in order for a candidate to be qualified for a position. Non-Competitive means there isn't one. The rest of the columns contain information that would be commonly shown in the job descriptions for most job seekers to apply.

As large as this dataset seems, there are some columns that have a considerable amount of missing data. 3983 of the listings have blank values under the "Post Until" column. 5790 of the listings have blank values under the "Recruitment Contact" column. 4056 of the listings have blank values under the "Work Location 1" column. 4384 of the listings have blank values under the "Hours/Shift" column. 2186 of the listings have blank values under the "Preferred Skills" column.

Data Cleaning and Preparation

We started out with removing the following columns we did not need for our analysis: "To Apply", "Post Until", "Title Code No", "Recruitment Contact", "Posting Updated", "Process Date", "Work Location 1", "Residency Requirement" and “Hours/Shift” These columns either had too many missing values or they did not contain any insightful data. After that, we simplified the Job Category column into the following categories because there were 186 unique categories:

Admin & HR, Building Ops, Comms & IG Affairs, Community Programs, Eng Arch Planning, Finance, Green Jobs, Health, Legal Affairs, Policy, Public Safety, Social Services, and Tech. To make it easier to graph data over time, we added a new column called “Posting Year” that only included the Posting Year for each listing.

Findings

From counting the job listings by Career Level, we can see that the dataset is mostly made up of experienced, non-manager roles with over 4200 listings, which is over 70% of the dataset. Student and Entry-Level jobs made up of 182 and 651 of the total listings. Most of the jobs from the dataset were posted in 2025 with most of them being experienced non-manager jobs, which makes sense. Almost exactly half of the dataset is made up of internal listings. More than 25% of the jobs posted are Engineering, Architecture and Planning roles. The Tech department has the highest average salary ($99,454), while the Building Ops department has the lowest average salary ($66494.87), according to the average salary by job category chart. The median minimum ($66430.0) and maximum salaries ($91768.0) fall into the same groups. Legal affairs has the highest maximum standard deviation ($43045.5), whereas Green jobs have the lowest ($26,405). Job listings for students are mostly in the Engineering, Architecture and Planning department, followed by Health and Community Programs. If university students are interested in working a city job, then those are the categories they should aim for. For most students, NYC jobs don’t seem to be a great place to look for listings simply because there just aren’t many jobs that are open to students.