Exploratory Data Analysis of NYC Government Job Postings

Authors:

Zaid Mohamad

Zhuowen Yan

Table of Contents

1. Abstract:

This EDA (Exploratory Data Analysis) project explores job postings available on the City of New York's official jobs site (http://www.nyc.gov/html/careers/html/search/search.shtml) and provides valuable insights for job seekers. By analysing over 5,000 postings from November 10 2020 to November 1 2024, the analysis focuses on trends in job categories, salary distributions, career levels, and in-demand skills. Data pre-processing included standardizing salary information to an annual basis and extracting temporal features for seasonal trend analysis. Using a combination of univariate, bivariate, multivariate, time and text-based analyses, the report highlight high-demand fields, salary patterns across career levels, and common qualifications needed. The analysis reveals that "Engineering, Architecture & Planning" is the most frequently posted job categories, whereas roles in "Technology, Data & Innovation," "Policy Research & Analysis," offer some of the highest starting salaries. Additionally, excellent verbal and written communication skills and proficiency in Microsoft Office are significant featured as essential qualifications. The data also shows that the number of job postings tends to be higher from September to early November, with October seeing the peak. These insights aim to help individuals make informed career decisions within the NYC job market.

2. Introduction:

The New York City job market has a wide range of career opportunities across different fields, especially in government agencies. For people looking to work in public service, understanding the high-demand fields, salary expectations, and required skills in specific job categories can be helpful. This project uses data from over 5,000 NYC government job postings to provide insights that can guide job seekers in making informed career choices.

The dataset includes details on job categories, career levels, salary ranges, qualifications, and preferred skills, focusing only on full-time positions, as part-time jobs represent just 0.2% of the dataset. By analyzing these factors, we can answer key questions like:

- **Which fields have the highest demand?** Identifying which job categories have the most postings, and shows which fields the city is actively recruiting.
- **What are the typical salary ranges?** Understanding salary distributions and comparing career levels helps job seekers know what to expect for different roles.
- **What qualifications and skills are commonly required?** By looking at the qualifications and skills in the postings, we can see the most valued abilities in NYC government jobs.
- **Are there any seasonal trends in hiring?** Analyzing timing helps us understand whether job postings peak at specific times, which may indicate potential hiring cycles.

To answer these questions, the analysis includes data pre-processing steps like filtering for full-time jobs, standardizing salary data to an annual basis for fair comparisons, and extracting time-based features to analyze trends. The analysis uses univariate and bivariate methods to explore individual data points and relationships between them. We also used text analysis to provide insights into frequently needed skills.

Through this analysis, the project aims to provide a comprehensive overview of the NYC government jobs. These insights can be especially useful for students, recent graduates, and professionals looking for public service roles, allowing them to better align their career planning and job search strategies with the current needs of the NYC job market.

3. Dataset Overview

This section provides an overview of the NYC Job Postings dataset, which is the foundation for our analysis. The dataset includes detailed information on job postings from NYC government agencies, providing insights into job categories, salary ranges, required qualifications, preferred skills, and other relevant details.

Source and Size

- **Source:** NYC Open Data Portal - https://catalog.data.gov/dataset/nyc-jobs
- **Dataset Size:** The dataset has 5,459 rows and 30 columns, with each row representing a unique job posting.

Key Attributes

- **Job Information**
  - **Agency:** Name of the City agency where the vacancy exists.
  - **Job Category:** The occupational group in which the posted job belongs.
  - **Full-Time/Part-Time Indicator:** Indicates if the job is full-time or part-time. Since part-time jobs only represent 0.2% of the dataset, they will be removed during the data pre-processing phase.
- **Salary Information**
  - **Salary Range From:** Proposed salary from: range. (Can be 0).
  - **Salary Range To**: Proposed salary to: range. (Can be 0).
  - **Salary Frequency:** Proposed salary frequency, such as Hourly, Monthly, Yearly, etc. This field needs to be standardized to make sure all salaries are on an annual basis.
- **Qualifications and Skills**
  - **Career Level:** Possible career level of the job (e.g., Entry-Level, Experienced, Manager).
  - **Minimum Qual Requirements:** Minimum qualifications required for position (e.g., degrees, years of experience).
  - **Preferred Skills:** Skilles preferred for this position, such as specific technical proficiencies or soft skills.
- **Time Information**
  - **Posting Date:** Date the job was posted, which helps to analyze trends over time.
  - **Post Until:** Date the posting will be removed.

4. Literature Review

The literature review provides context by exploring past research on job market analysis, skill trends, and career planning. It helps place this project within existing research and shows how data analysis can support job seekers in making well-informed decisions.

Job Market Trends and Data Analysis

Previous studies have shown that using data to understand the job market can be very helpful for job seekers. It can show industry demand, needed skills, and salary expectations. For example, research from the **Bureau of Labor Statistics (BLS)** regularly examines trends in jobs and which industries are expected to grow. This helps job seekers plan their careers with fields that are in high demand [1].

Skills Analysis in Job Postings

Research from **LinkedIn Economic Graph** has used job posting data to find out which skills are in the most demand. This helps education providers and job seekers understand the changing landscape of skill needs [2]. This approach matches ours, where we use text analysis to find frequently mentioned qualifications and skills in NYC government job postings.

Salary Analysis and Career Levels

Research has found that government jobs often have specific salary ranges based on job level and standardized pay scales, which makes public sector job data useful for comparing salaries. Studies in both public and private sectors show that higher career levels, like managerial and executive jobs, usually offer higher starting salaries. In this project, we also look at salary ranges across different job levels in NYC government positions.

Seasonal Trends in Hiring

Seasonal trends in hiring have been explored in research on labor economics [1]. These studies suggest that job availability goes up and down throughout the year. Many organizations, including government agencies, increase hiring during specific times, like the start of a fiscal year or after budgets are approved. Our time-series analysis looks at similar seasonal trends in NYC government hiring to guide job seekers on the best times to apply.

Applying Data-Driven Insights to Career Planning

Data-based career planning is becoming more common because job seekers want to make good decisions based on real-time job market data. Websites like **Indeed** and **Glassdoor** use their user data and job postings to provide advice on careers, salary comparisons, and skill needs. This project takes a similar approach, using **Exploratory Data Analysis (EDA)** to find practical insights from NYC government job postings. These findings can help people understand high-demand fields, needed skills, and possible salary expectations, which will allow for better career planning.

5. Architecture/Methodology

This section explains the tools, methods used to do the Exploratory Data Analysis (EDA) on the NYC Job Postings dataset. The goal is to extract valuable insights that help job seekers make informed decisions.

Methodology Overview

The project uses a structured methodology that includes data cleaning, transformation, analysis, and visualization. Each stage is important for the accuracy and depth of the insights from the data.

1. Tools Used

- **Python:** The main programming language used for data analysis and visualization.
- **Pandas:** A Python library used to handle, clean, and prepare the dataset.
- **NumPy:** Used for mathematical operations like calculating Z-scores to find outliers.
- **Matplotlib and Seaborn:** Libraries for making charts and graphs that help visually explore the data.
- **Scikit-Learn:** Specifically, **CountVectorizer** was used for text analysis to identify common skills and qualifications.

2. Data Pre-processing and Cleaning

- **Handling Missing Values:** Missing values in key columns like **Minimum Qualifications** and **Preferred Skills** were filled with empty strings to avoid null values and make the text analysis work.
- **Filtering Out Part-Time Jobs**: Since part-time jobs were only 0.2% of the dataset, they were removed. Focusing only on full-time jobs helped to make the analysis more clear and relevant.
- **Standardizing Salary Information:** Salaries listed as "Hourly" "Monthly" or "Yearly" were converted to an annual basis for uniform comparison.
- **Outlier Detection and Removal:** Outliers in **Salary Range From** and **Salary Range To** were found using Z-scores. Rows with Z-scores above 3 were removed to focus on realistic salary values.
- **Date Transformation:** The **Posting Date** column was used to create **Posting Year** and **Posting Month** columns to make it easier **[changed from "facilitate"]** to analyze trends.
- **Duplicate Removal and Column Dropping:** Duplicate rows were removed, and unnecessary columns were dropped to create a cleaner, more focused dataset.

3. Data Analysis Techniques

- **Univariate Analysis:** Looked at individual variables and their distributions:
  - Analyzing job categories to find out which fields had the highest number of postings.
  - Investigating salary distributions to understand typical pay.

- o Looking at career levels to see the breakdown of entry-level, experienced, and manager roles.
- **Bivariate Analysis:** Used to explore how two variables relate to each other:
  - o Comparing starting salary by career level to see differences in pay.Maximum salary by career level to understand the salary ceiling for each level.
  - o Analyzing maximum salary by career level to find the salary limit for each level.
  - o Examining starting salaries by job category to identify higher-paying fields.
- **Multivariate Analysis:** Looked at relationships between career level, Salary Range From, and Salary Range To. The correlation heatmap helped identify significant relationships between variables like career level and salary range.
- **Time Series Analysis:** Analyzed trends over time by using Posting Year and Posting Month to find hiring peaks and possible seasonal patterns in job postings.
- **Text Analysis:** Used **CountVectorizer** from **Scikit-Learn** to find common skills and qualifications mentioned often in postings. This gave insights into the most valued skills for NYC government jobs.

4. Data Visualization

Visualizations played a key role in making the data easier to understand. The main types of visualizations used were:

- **Bar Charts:** Showed the distribution of job categories, top skills, and salary ranges by job category.
- **Box Plots and Violin Plots:** Showed salary distributions by career level and job category, highlighting differences in pay.
- **Heatmaps:** Used to show correlations between **Career Level, Salary Range From,** and **Salary Range To**, helping to find important relationships.
- **Line Plots:** Illustrated time trends in job postings by month and year.

5. Steps of Analysis The structured EDA process involves the following steps:

- Load and Inspect Data
- Preprocess Data
- Univariate Analysis
- Bivariate Analysis
- Time Series Analysis
- Text Analysis on Skills and Qualifications
- Visualization

The project uses a step-by-step approach to clean, transform, and analyze the NYC Job Postings dataset. By combining data preparation, statistical analysis, and visualizations, the analysis provides a comprehensive view of NYC government jobs, focusing on job categories, pay, career levels, and required skills. The goal of these insights is to help job seekers, especially those interested in public service, make better career decisions.

6. Data Cleaning

The first step in our analysis was **data cleaning**. This was crucial for making sure our dataset was free of errors, any kind of irrelevant information, and missing data that could interfere with our insights. The following steps were carried out:

1. Removing Duplicate Rows:

Duplicate rows were found and removed to make sure each job posting was unique. Removing duplicates cuts down on unnecessary repetition and helps prevent skewed results in the analysis.

2. **Dropping Irrelevant Columns:**

Some columns, such as **Recruitment Contact, Division/Work Unit, To Apply**, and **Post Until**, were not relevent to the goals of our analysis and were removed. Removing these columns reduced unnecessary data and made the dataset clearer by keeping only what was necessary for the analysis.

3. Handling Missing Values:

Columns like **Minimum Qualifications** and **Preferred Skills** had missing values, which were filled with empty strings. This was done to avoid having null values that could cause problems in text-based analysis methods like word clouds or frequency counts.

7. Pre-processing

Data pre-processing involved several important steps to standardize the data, making sure it was accurate and consistent for analysis. Each step addressed specific problems in the dataset, such as inconsistent salary formats and outliers.

1. Standardizing Salary Information:

The **Salary Frequency** column had values like **"Hourly", "Monthly",** and **"Yearly"** which made comparing salaries difficult. To solve this, all salary values were converted to an annual basis:

- **Hourly rates** were multiplied by 2,080 (the typical number of working hours in a year).
- **Monthly rates** were multiplied by 12 (months in a year).
- **Yearly rates** were left unchanged.

This change makes all salary data easy to compare, no matter what the original frequency was.

2. **Filter for Full-Time Jobs**:

Focus only on full-time job opportunities. So, all part-time listings were filtered out to give a clearer picture for those seeking stable, long-term careers.

3. Remove "Student" Level Jobs:

Job listings that were specifically for students were removed. This allowed us to concentrate on professional roles, which are more relevant to career seekers.

4. Removing Outliers:

Outliers in the **Salary Range From** and **Salary Range To** columns were found using the **Z-score** method. Rows with Z-scores higher than 3 were treated as outliers and removed. This step focused the salary data on realistic values and prevented extreme numbers from affecting the analysis too much.

8. Feature Engineering

1. **Date Transformation:**

To help with analyzing trends, the **Posting Date** column was used to create new columns called **Posting Year** and **Posting Month**. These new columns make it easier to analyze when job postings happen over time, making it possible to spot any seasonal patterns in hiring.

2. **Creating Salary Bins:**

To make it easier to analyze salary ranges, salary bins were created using the **Salary Range From** values. The dataset was divided into four bins: **Low (0–50,000)**, **Medium (50,000–100,000)**, **High (100,000–150,000)**, and **Very High (150,000–300,000)**. Creating these bins helps with straightforward comparisons between different salary levels and shows trends in each category.

3. Convert Career Levels to Numeric Values: To help with further analysis, we converted career levels into numbers:

- **Entry-Level**: 1
- **Experienced**: 2
- **Manager**: 3
- **Executive**: 4

This conversion made it possible to use career level data in mathematical calculations, such as correlations.

4. Simplify Job Categories:

We noticed that many job categories had sub categories, we replaced all the sub categories with the main category for effective analysis

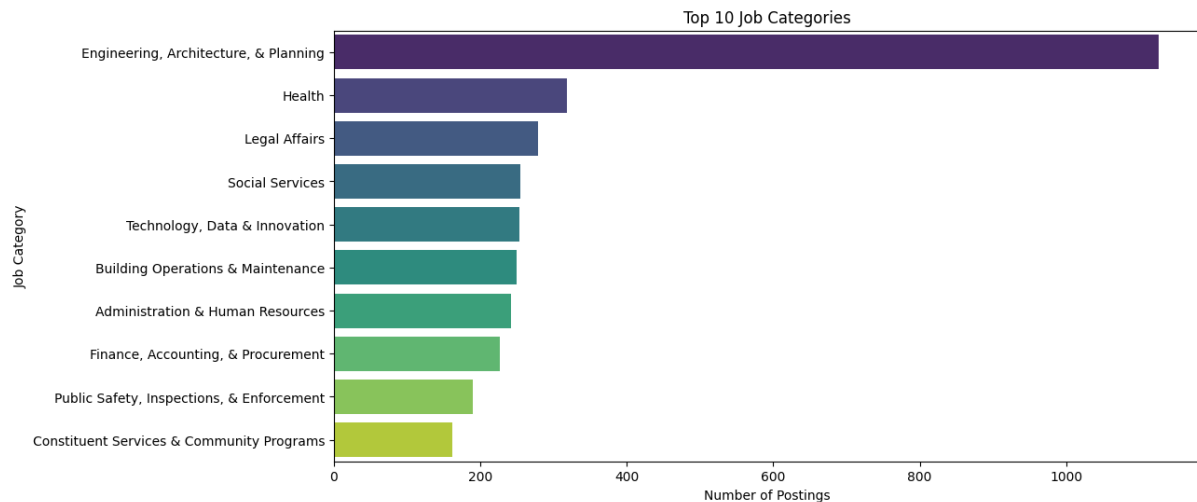# 9. Data Visualization and Insights

9.1 Univariate Analysis

Figure 1.1 Top 10 Job Categories
The distribution of job categories shows that the most in-demand field is **Engineering, Architecture & Planning**. Engineering roles are particularly common, highlighting a high demand for technical skills in city projects and infrastructure. Other important fields like **Health, Legal Affairs,** and **Social Services** show the city's need for professionals to support public health and welfare.
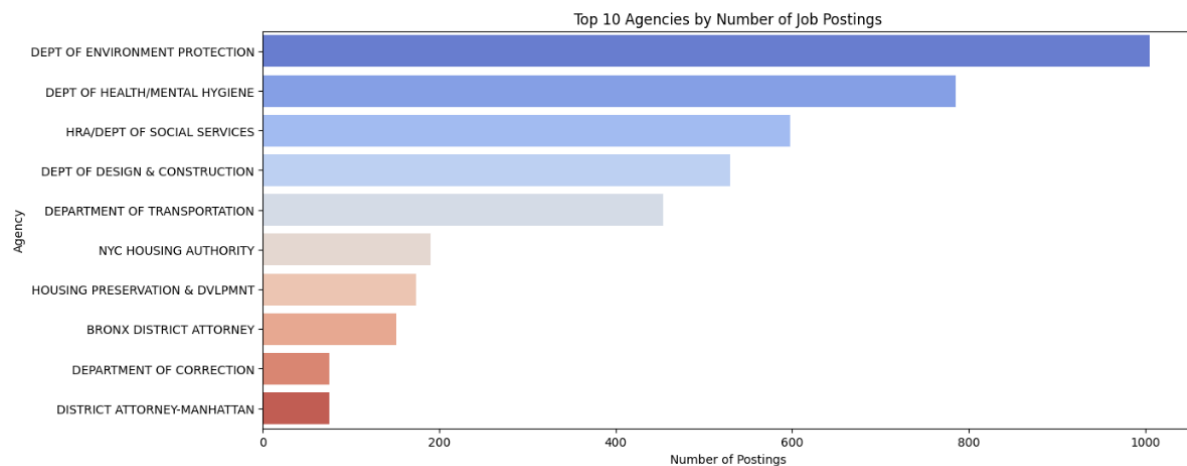
Figure 1.2. Top 10 Agencies
The analysis of job postings by agency shows hiring patterns, with the **Dept of Environmental Protection**, **Dept of Health/Mental Hygiene**, **HRA/Department of Social Services**, **Dept of Designing & Construction**, and **Dept of Transportation** having the most open positions. This means these agencies are key employers in the NYC public sector, likely because of their essential services that need continuous staffing.

Distribution of Career Levels
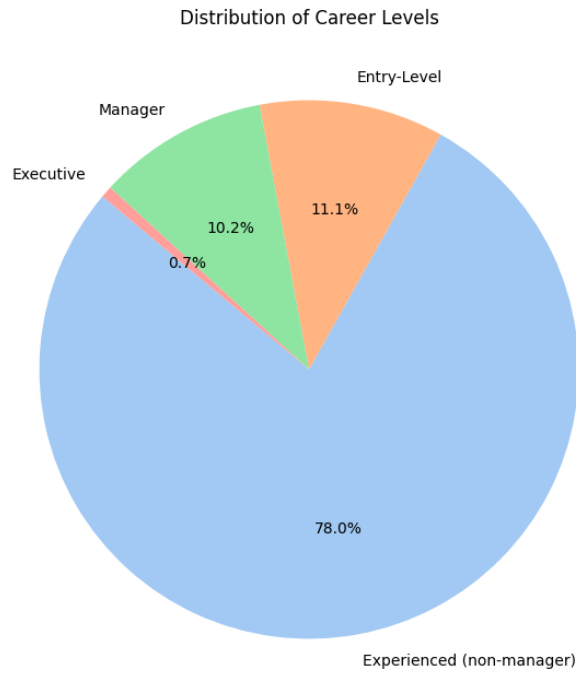
Figure 1.3. Distribution of Career Levels
The career level distribution shows that most jobs are for **Experienced (non-manager)** roles**,** which make up 78% of the listings. This is followed by **Entry-level** and **Manager** positions, with executive roles being the least common. The high number of experienced roles suggests that NYC agencies prefer people with relevant past experience, likely due to the demands of public service jobs.
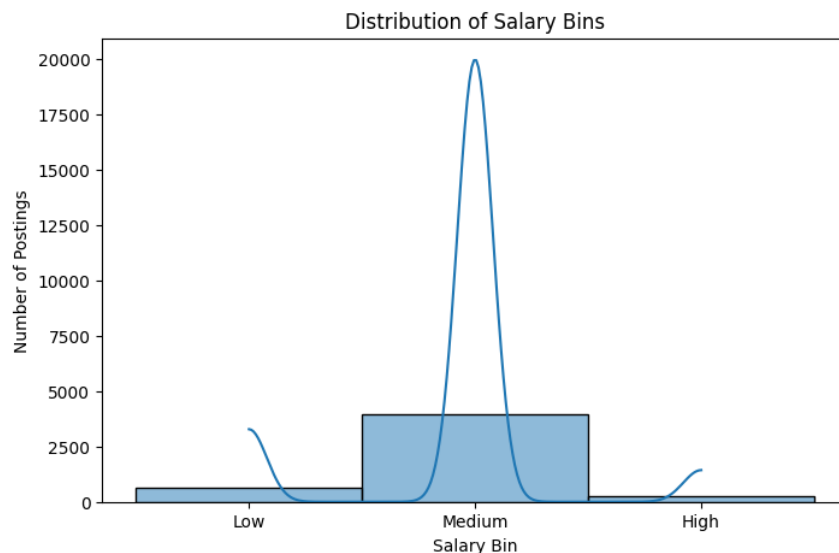

Distribution of Salary Bins

Figure 1.4. Distribution of Salary Bins
The majority of jobs fall into the **Medium** salary range ($50,000 to $100,000 annually). This suggests that NYC government jobs generally offer moderate pay.
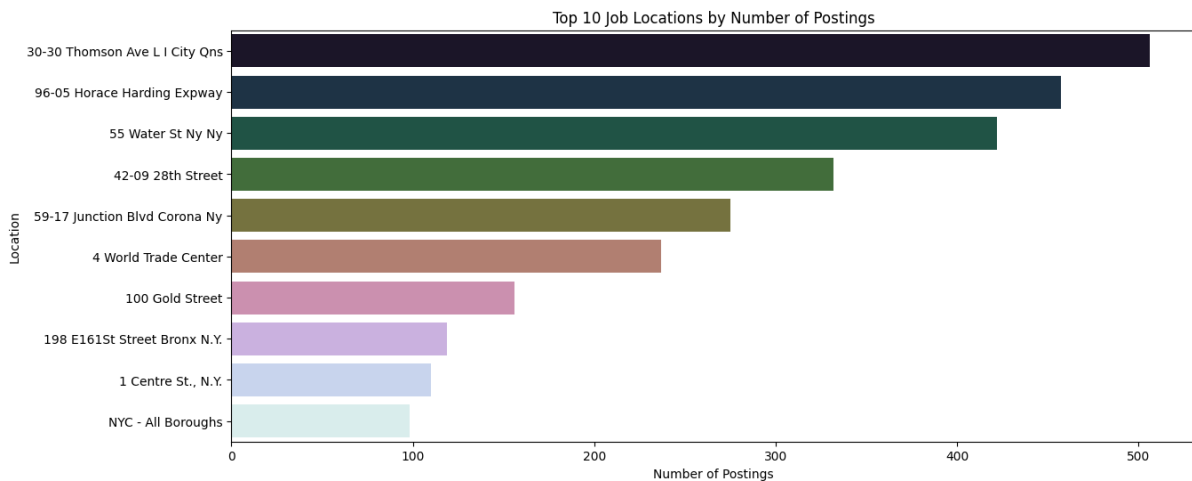
Figure 1.5. Top 10 Job Locations.

The analysis also considers job location distribution, with certain locations, such as **Thomson Ave, Horace Harding Expressway,** and **Water St.**, being major hubs for job postings. This concentration of job postings suggests that specific areas may serve as primary centers for public sector employment in NYC.
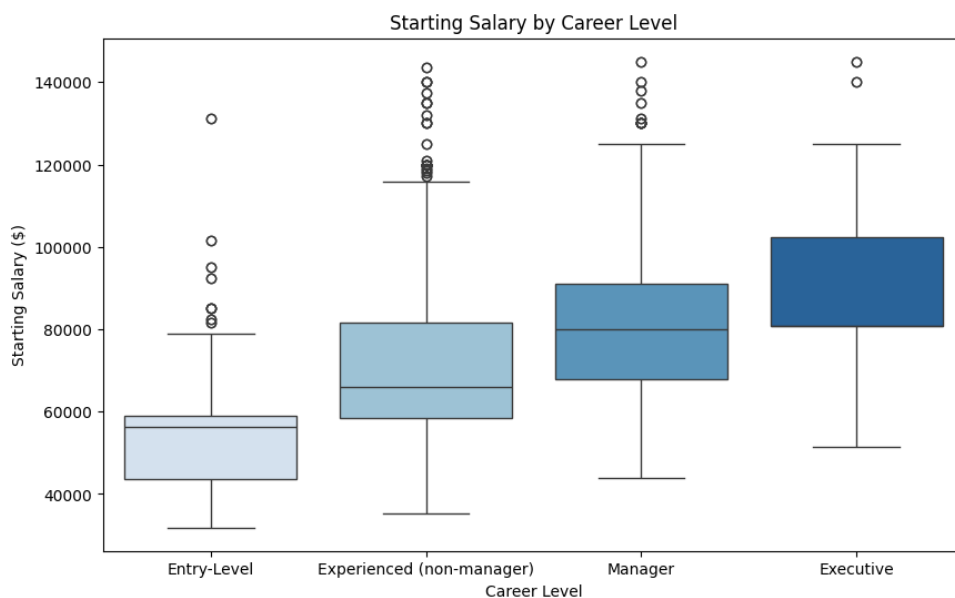
9.2. Bivariate Analysis

Figure 2.1. Starting Salary by Career Level
An analysis of starting salaries by career level shows that, as expected, higher career levels come with higher starting salaries. Executive roles offer the highest starting pay, followed by managerial roles. This trend matches industry norms, where senior roles get higher pay because of their added responsibilities.
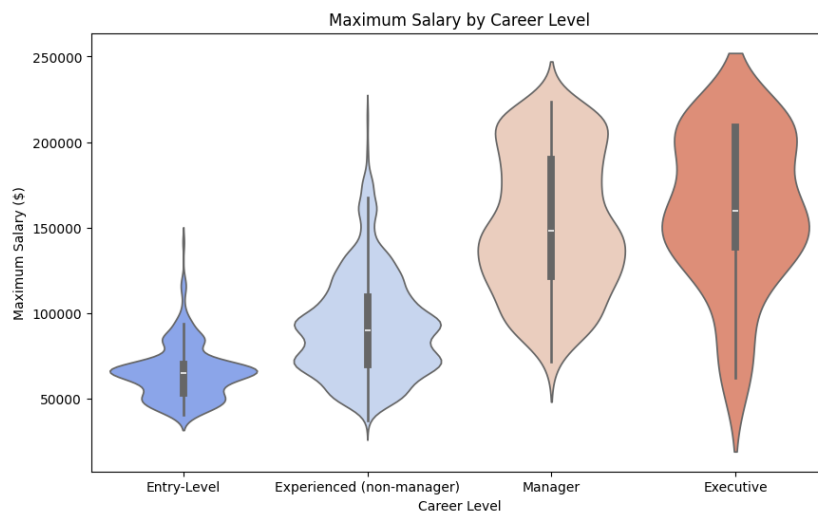
Figure 2.2. Maximum Salary by Career Level
A review of the maximum salary distributions shows that executive roles have the highest earning potential, with a wide range of salaries, while entry-level roles have a smaller salary range, showing limited earning potential.
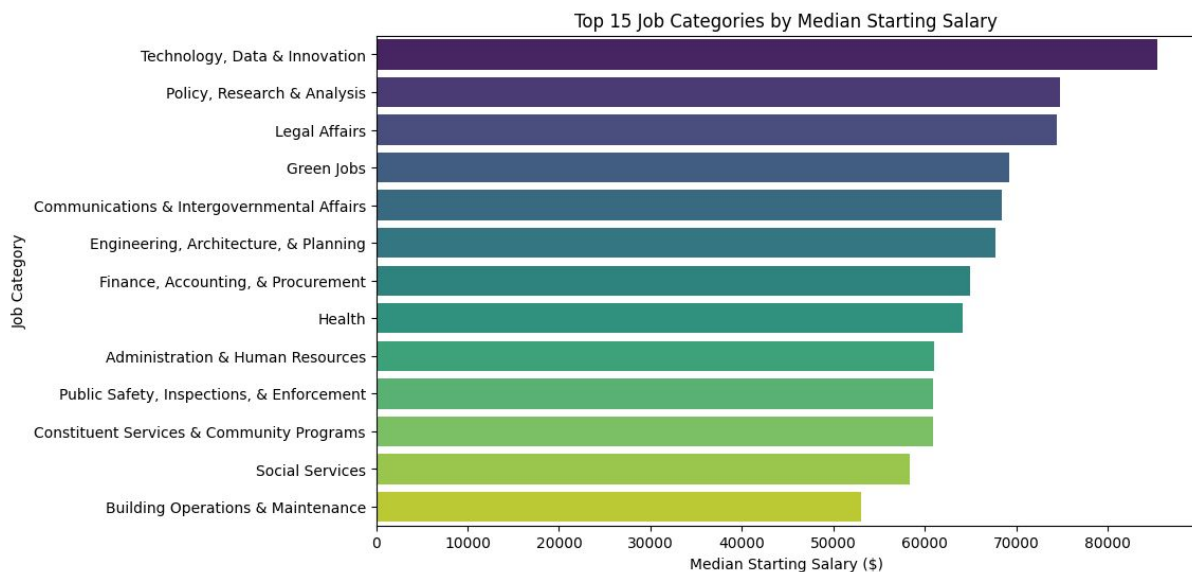

Figure 2.3. Top 10 Job Categories by Median Starting Salary
This figure ranks the job categories based on the median starting salary, helping job seekers understand which fields are more financially rewarding. The analysis indicates that job categories involving **technology, research,** and **legal expertise** are the most financially rewarding, with starting salaries well above the median for other public sector roles. Conversely, categories related to **social services** and **operations** tend to offer lower starting salaries.
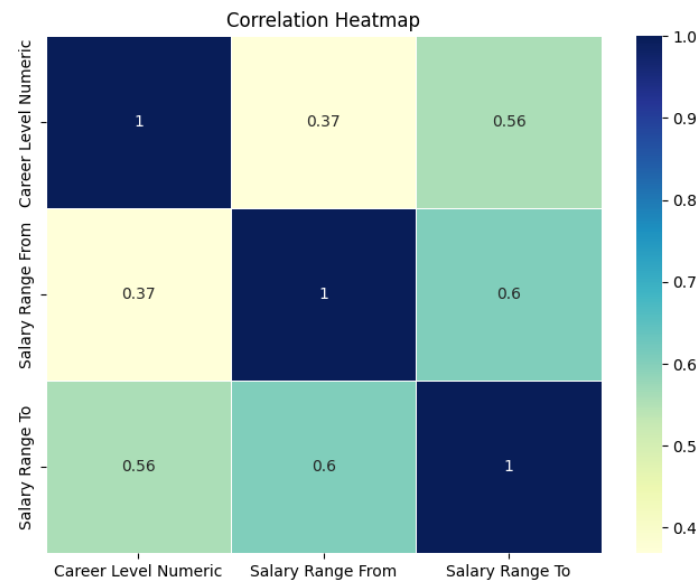
9.3. Multivariate Analysis

Figure 3.1. Correlation Heatmap

A correlation heatmap was used to assess relationships between career level, starting salary, and maximum salary. The results show a moderate positive correlation between career level and both starting and maximum salaries. This reinforces the idea that higher career levels lead to higher pay in NYC government jobs.
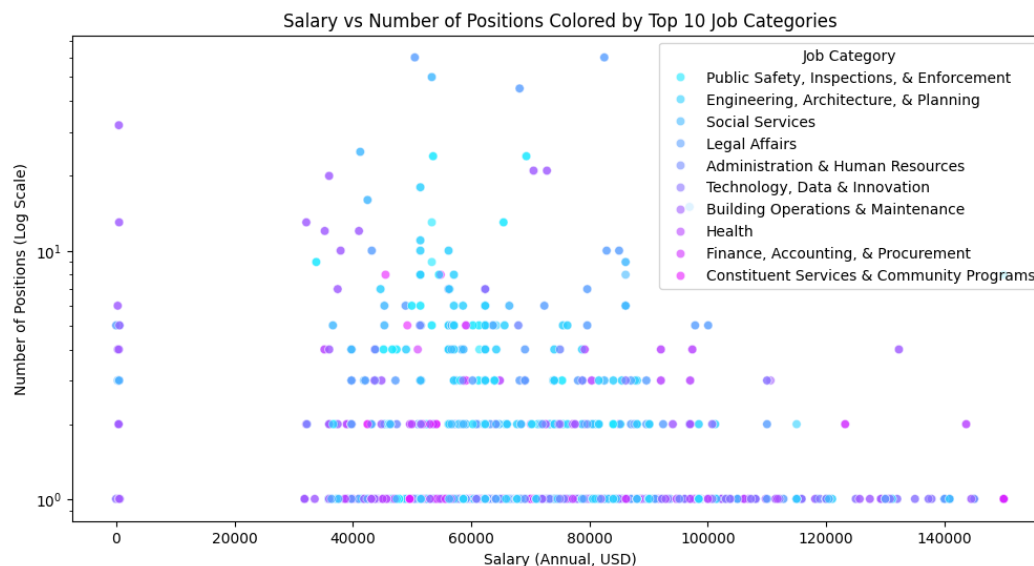

Figure 3.2. Salary vs Number of Jobs in Top 10 Job Categories.

This analysis shows a trade-off between high salaries and job availability in NYC government jobs. Fields like **Public Safety** have numerous openings but generally offer moderate pay, whereas fields such as **Legal Affairs** provide higher salaries but fewer opportunities. Job seekers may need to balance their preferences between higher compensation and greater job availability depending on their career goals in the public sector.
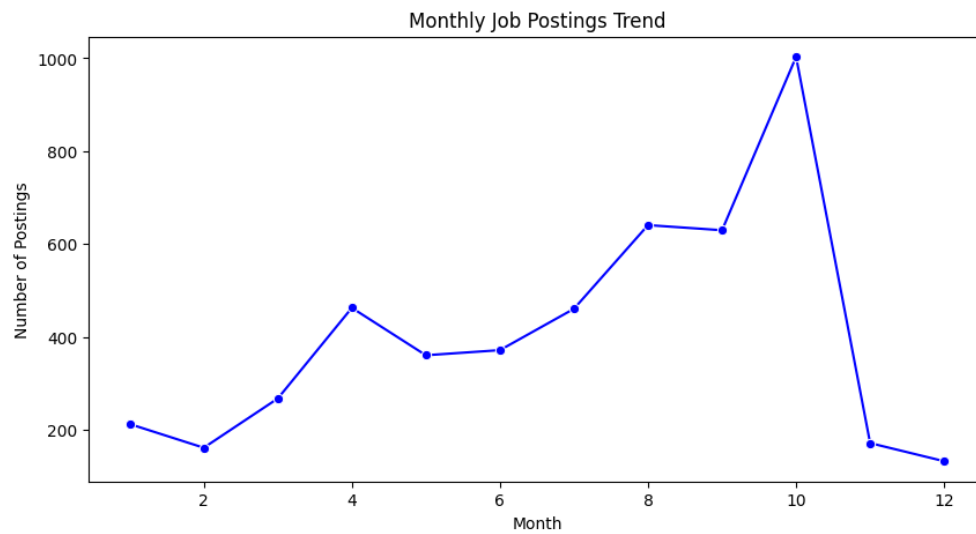
9.4. Temporal Analysis

Figure 4.1. Monthly Job Postings Trend

The monthly trend of job postings shows seasonal patterns, with a noticeable peak in **October**. This suggests that hiring may be affected by the fiscal year or budget cycles, which can help job seekers understand the best times to apply.
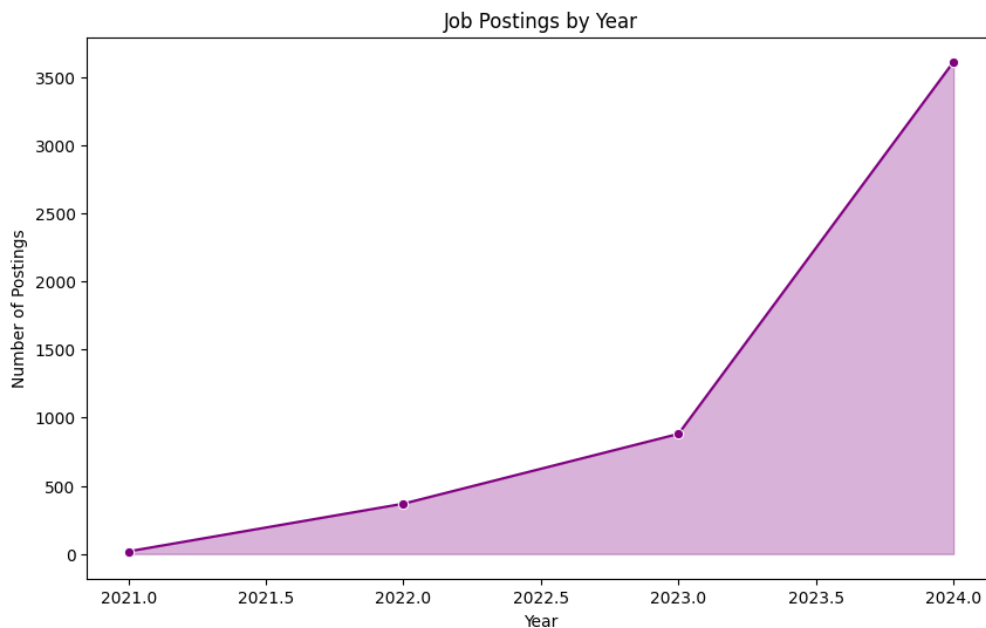

Figure 4.2. Job Postings by Year

The yearly trend analysis shows a sharp rise in job postings in **2024**, likely reflecting more hiring or new projects. This suggests strong demand for government jobs in recent years.
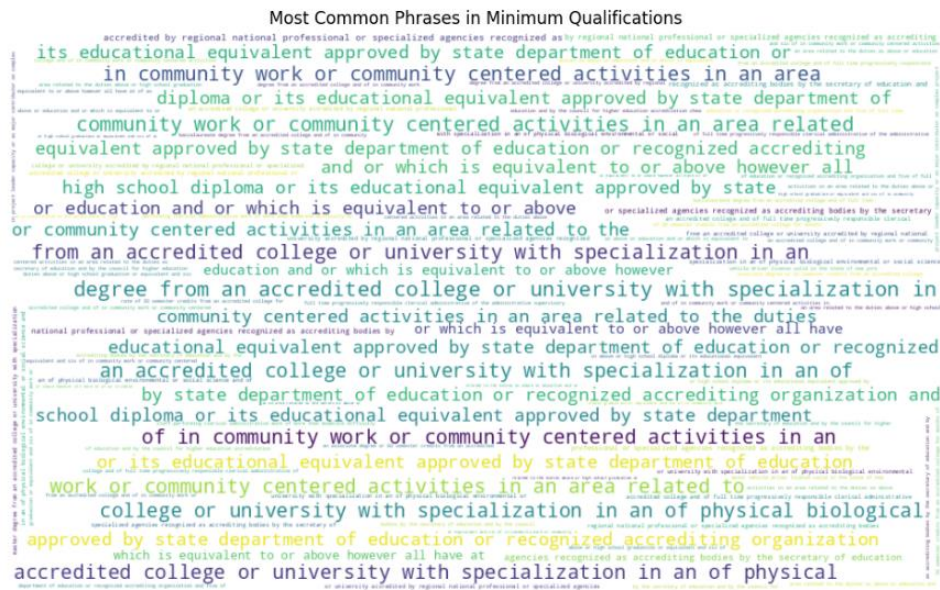
9.5. Text Analysis

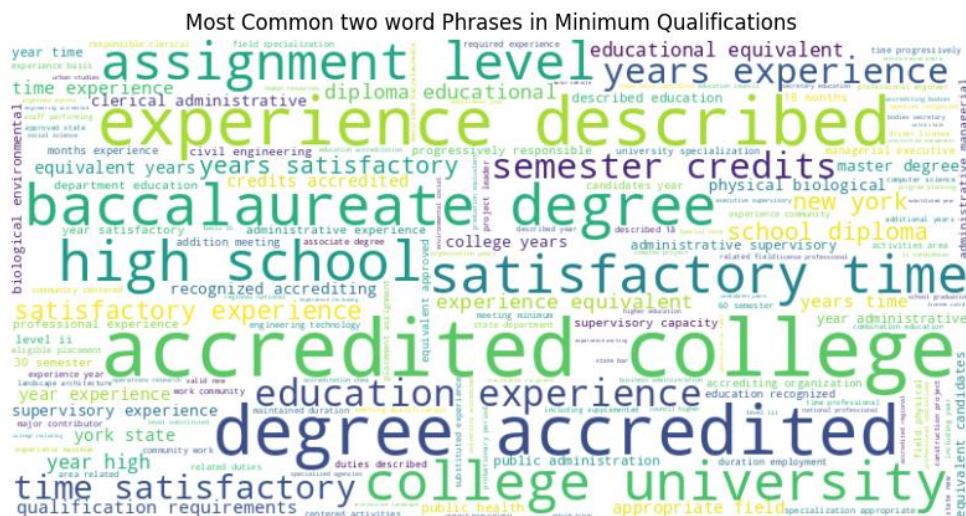**Figure 5.1.a. Most Common 8-word Phrases in Minimum Qualifications**

Figure 5.1.b. Most Common two-word Phrases in Minimum Qualifications

An analysis of two-word phrases and 8-word phrases in **Minimum Qualifications** shows that many NYC government jobs need formal education, with phrases like "**Four year diploma**," "**accredited college**," and "**satisfactory experience**" appearing often. A word cloud shows this emphasis on education and prior experience.
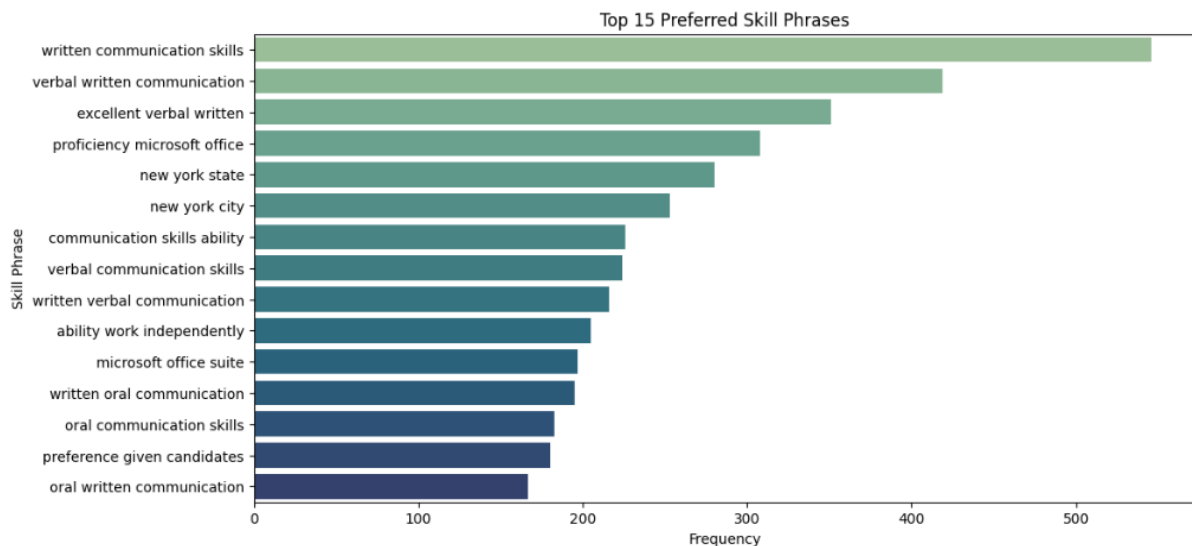
Top 15 Preferred Skill Phrases

Figure 5.2. Top 15 Preferred Skill Phrases

The top skills highlighted in the analysis are "**written communication skills**", "**verbal written communication**" "**excellent verbal and written**" and "**proficiency in Microsoft Office**." These skills are commonly requested across many roles, emphasizing the importance of effective communication and digital proficiency for NYC government jobs.

10. Conclusion:

- **High Demand Fields:** NYC government job postings show significant demand in **Engineering, Health, Legal Affairs**, and **Social Services**, which are essential for city infrastructure and public services.
- **Key Hiring Agencies:** The **Department of Environmental Protection** and **Department of Health/Mental Hygiene** lead in hiring, indicating a focus on environmental and health services.
- **Experience Preferred:** Most jobs are for **experienced non-manager**, with fewer entry-level positions, showing a preference for candidates with relevant experience.
- **Salary Patterns:** Pay generally increases with career level, though most jobs offer moderate salaries in the **$50,000 to $100,000** range.
- **Skill Requirements:** The most desired qualifications are educational degrees and communication skills, with **Microsoft Office** proficiency and strong interpersonal abilities often listed.
- **Hiring Trends:** Job postings are highest in **October**, possibly related to fiscal cycles, and show a steady increase in recent years, which likely reflects more hiring**.**
- **Guidance for Job Seekers**: This analysis helps job seekers in aligning their skills and timing with public sector opportunities.

11. Recommendations and Future Directions

- **Enhanced Skill Matching for Job Seekers:** Since communication and Microsoft Office skills are in high demand, future analysis could explore specific skill gaps by examining the supply of these skills among applicants. Partnerships with educational institutions to provide targeted training programs could help better align the workforce with public sector needs.
- **Seasonal Hiring Patterns:** Since **October** had the highest number of job postings, more analysis could explore seasonal or cyclical hiring trends over several years. This would help job seekers better plan their applications.
- **Predictive Hiring Needs:** By using historical data, predictive models could be created to estimate future hiring needs across NYC agencies. This would help policymakers prepare in advance by aligning resources and recruitment efforts to meet upcoming demands.
- **Analysis of Job Flexibility and Benefits:** Future analysis could explore the flexibility options and additional benefits offered by NYC government jobs, such as remote work opportunities, health benefits, and leave policies. Understanding these aspects could help job seekers make informed decisions that go beyond salary considerations and align their career choices with their personal needs and lifestyle preferences.

These recommendations are intended to build on the current analysis, providing deeper insights that can help shape job-seeking strategies and workforce planning in NYC's public sector.

12. References:

1. "The Employment Situation", News Release, Bureau of Labor Statistics,U.S. Department of Labor, November 1, 2024.

2. "Skills-First: Reimagining the labor Market and Breaking Down Barriers", LinkedIn Economic Graph, 2023.

3. https://www.kaggle.com/code/imoore/intro-to-exploratory-data-analysis-eda-in-python

4. https://www.geeksforgeeks.org/exploratory-data-analysis-in-python/