JOB RECOMMENDATION SYSTEM

PROJECT PRESENTED By : GROUP ONE

- LOYCE TSUMA
- HANSEL OMONDI
- TATIANA TEMBA
- AUSTIN MURUNGA
- PHIL KONRAD

CONTENTS

1 PROJECT OVERVIEW

This project aims to build a job recommendation system to help users find relevant job opportunities based on their interests, and preferences.

The Goal is to Improve job search efficiency by matching candidates with roles that align with their career goals and skills.

2 BUSINESS UNDERSTANDING

Problem: Many job seekers face difficulties finding relevant jobs due to the vast number of listings. This project aims to streamline the job search process by recommending similar jobs.

By leveraging job details like titles and descriptions, users are presented with a list of related opportunities to enhance their job search experience.

3 DATA UNDERSTANDING AND ANALYSIS

Data Overview

The dataset used in this project is the Combined Jobs Final dataset sourced from Kaggle, which contains various job listings including details such as job title, company, location, job description, and more.

The composition of the dataset is as follows

- Rows: 84,090
- Columns: 23

3 DATA UNDERSTANDING AND ANALYSIS

Main Columns

- **Job.ID**: Unique identifier for each job listing.
- **Title**: Job title or role.
- **Company**: Name of the hiring company.
- **City**: City where the job is located.
- **State.Name**: Name of the state where the job is located.
- **Industry**: Industry related to the job.
- **Job.Description**: Detailed description of the job role.
- **Requirements**: Qualifications and skills required.
- **Employment.Type**: Type of employment (e.g., full-time, part-time).

3 DATA UNDERSTANDING AND ANALYSIS

DATA PREPARATION

After assessing the data by column and checking for missing values, the following columns were dropped given their high percentage of missing values.

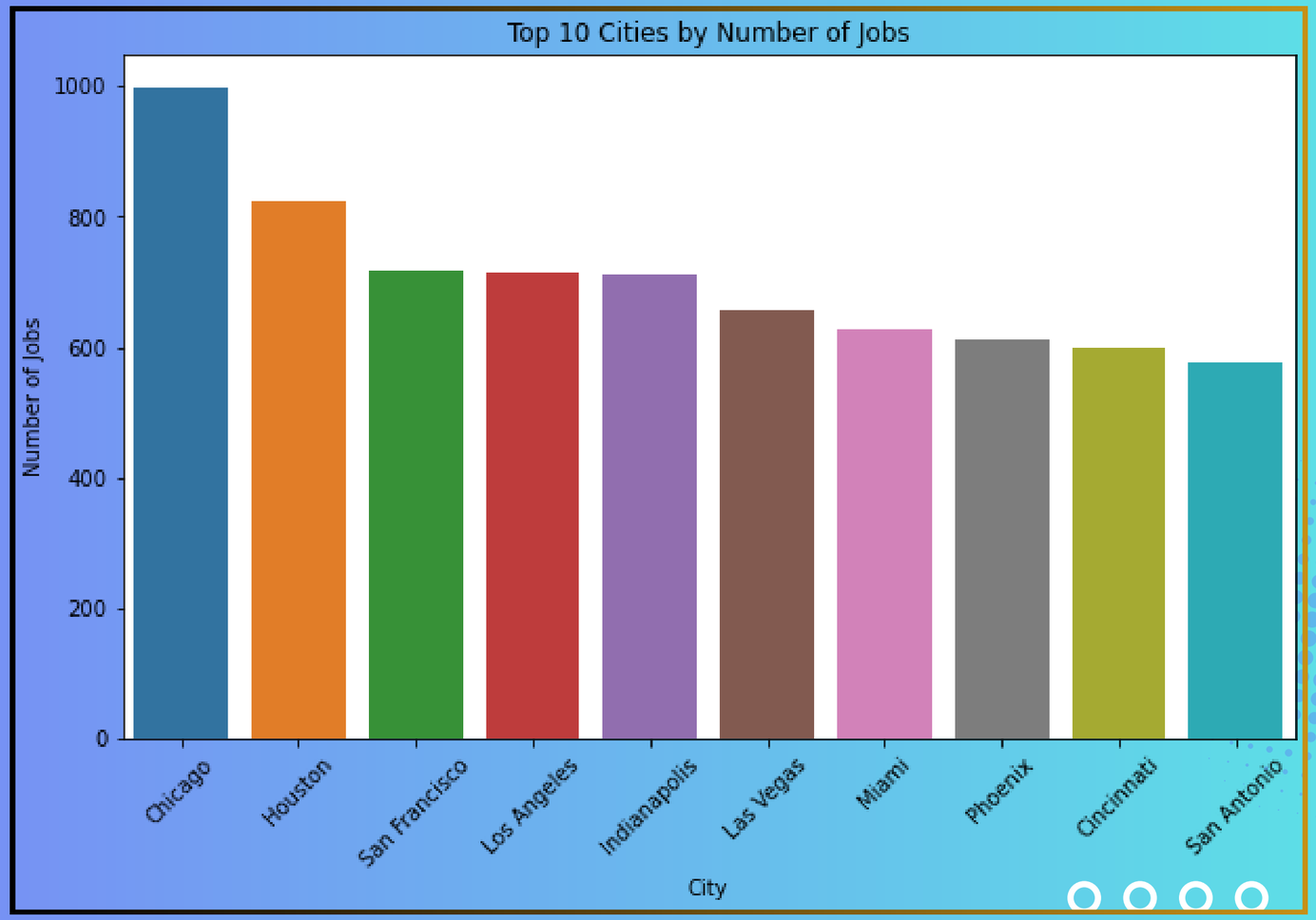
Requirements 100.000000 %
Address 99.957189%
Salary 99.727673%
Industry 99.682483%

Columns with unimportant information were also dropped i.e 'Listing.End', 'Listing.Start', 'Latitude', 'Longitude', 'State.Code', 'Status', 'Title', 'Slug'

3 DATA UNDERSTANDING AND ANALYSIS

EXPLORATORY DATA ANALYSIS

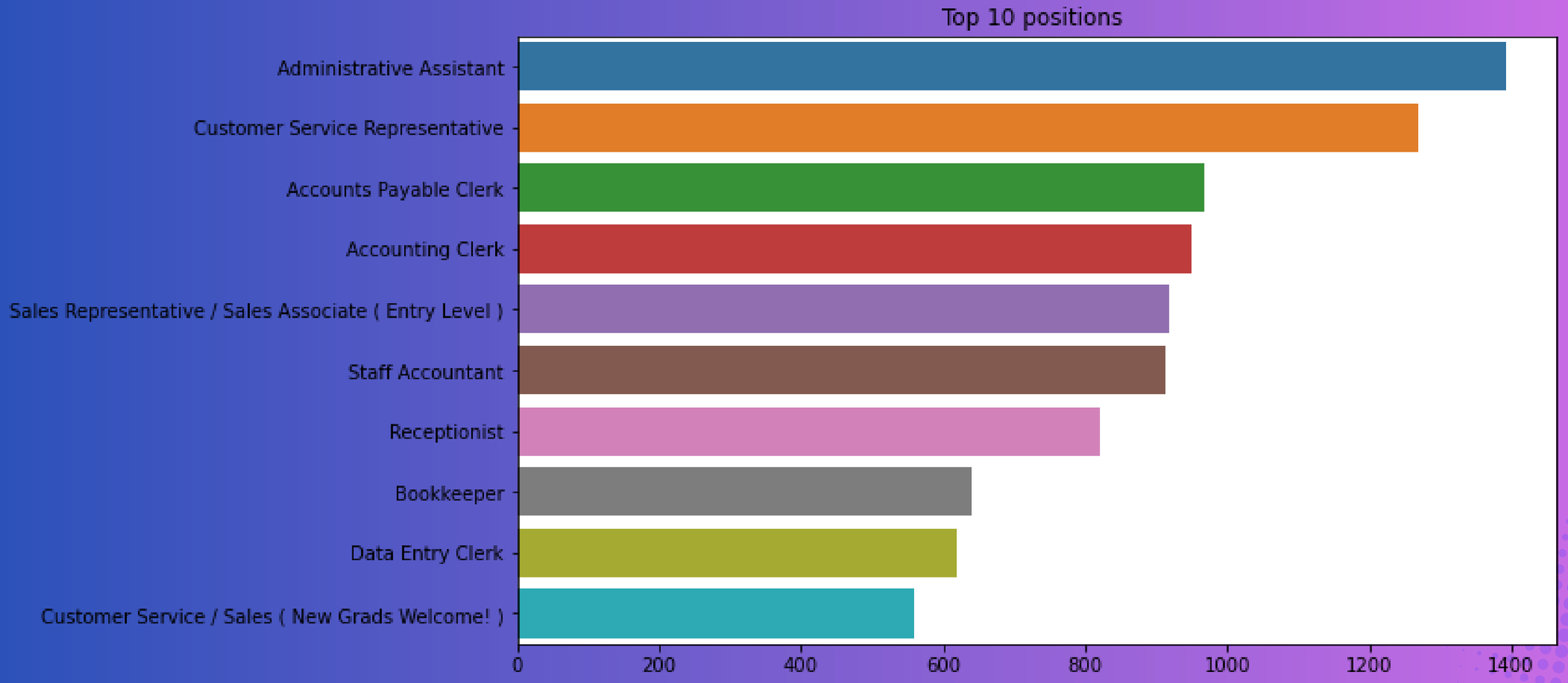
A univariate analysis was conducted and visualizations carried out and these were the results:



From this visualization we are able to see the top 3 cities with job listings are Chicago, Houston and San Francisco

3 DATA UNDERSTANDING AND ANALYSIS

EXPLORATORY DATA ANALYSIS



From the above image we can conclude that the top position is Administrative Assistant

4 MODEL DEVELOPMENT AND EVALUATION

Modeling Approach:

- The TF-IDF Vectorizer was used to convert job descriptions into numerical form.
- A K-Nearest Neighbors (KNN) algorithm with cosine similarity was employed to find similar jobs.

EVALUATION:

- The model was tested by providing job recommendations and evaluated based on user feedback and job relevance.
- Precision: The model achieved a high precision of 1.0, indicating all recommended jobs were relevant.
- Recall: Low recall (0.08), meaning the model could find only a fraction of all possible relevant jobs.

5 OPTIMIZATION STRATEGY

Improvement Focus: The recall rate can be enhanced by:

- Tuning hyperparameters like the number of neighbors.
- Exploring additional features beyond job descriptions to capture more diversity in the recommendations.
- Increasing the range of jobs considered in the recommendation set.

5 OPTIMIZATION STRATEGY

After the model hyper tuning, these were the results :

Precision = 0.4: A precision of 0.4 means that 40% of the jobs recommended were relevant (i.e., had the same job position as the original job).

Recall = 0.0339 (or about 3.39%): Recall represents the ratio of correctly recommended jobs to the total number of relevant jobs in the entire dataset. A recall of 0.0339 means that only about 3.39% of the total relevant jobs were recommended.

The recall is quite low signifying that the model is not covering all possible relevant jobs well.

6

INSIGHTS AND RECOMMENDATIONS

- The model provides highly relevant job recommendations but misses many other potential matches.
- Implement further optimizations to improve the recall.
- Expand the feature set to include factors like job requirements or salaries for more efficient recommendations.

GROUP ONE Presentation

THANK YOU