HDSC Winter '22 Premier Project

A Project by Team Gradboost

Scope: Jobs and Careers

Topic: Online Job Postings

INTRODUCTION

The demand for jobs in the labor market is continuously evolving with the use of technology to advertise job openings, hence, the need to understand the demand for certain job titles and professions. There is also a need to identify skills that are most frequently required by employers, and how the distribution of these skills changes over time, which can help in making recommendations to job seekers and employers.

The online job market is a strong predictor of overall labor needs in an economy. Furthermore, data from online job posts is easier and faster to obtain, and they can be a more comprehensive source of information than traditional job postings as seen in newspapers and print media.

In recent times, the job market has made significant improvements in the areas of advertisement. With the advent of social media and job/career recruitment sites, this has made for easy access to a collection of available jobs in the labor market. However, the challenge experienced by potential employees and recruiting employers is how to match appropriate job postings with qualified or eligible candidates.

As a result, this project sets out to obtain insights from analyzing an online job posting dataset and make inferences regarding the most in-demand job titles, varying skills in demand, and proffering recommendations to help job seekers and recruiting firms.

PROBLEM STATEMENT

This project is aimed at answering the following questions:

1. What jobs are in high demand?

2. Which skills are mostly required by employers?

3. How has the need for these skills changed over time?

ONLINE JOB POSTING DATASET

The dataset used for this project was obtained from a Kaggle notebook. The data was originally scraped from a Yahoo! Mailing group. Each row represents certain aspects of a job post. The dataset representation is tabular, but many of the columns are textual/unstructured in nature. Most notably, the columns JobDescription, JobRequirement, RequiredQual, ApplicationP, and AboutC are textual. The jobpost column is an amalgamation of these various textual columns. The dataset contains 19,001 job postings from 2004 to 2015 posted on Career Center, an Armenian human resource portal. This data described the nature of the ever-changing job market and results obtained from the analysis will help provide insights into the overall demand for labor in the economy.

DATA EXPLORATION

The dataset was downloaded from here. There were 24 columns containing the job posts and various aspects of the job post.

Key Observations:

1. Data entries in the jobpost column were scraped from the web and were processed to create other columns.

2. There were repeated entries of similar job posts in the 'jobpost' column, consequently, job posts from a company posted in the same year and month were duplicated.

3. The dataset had a large number of null values (30%).

4. The only significant columns to answer the objectives were columns with the required information such as 'Title', 'IT', 'Company', 'JobDescription, 'JobRequirment', 'RequiredQual', 'Month', and 'Year.

5. The dataset used in this project like every real-life data required cleaning for null values, duplicated observations, and redundant columns/rows.

DATA CLEANING

The following steps were taken to clean the dataset:

1. Columns in the dataset were renamed for easy identification.

2. All duplicated rows filtered with the jobpost, Title, Company, Month, and Year columns were removed.

3. Null values in Title, Job_Description, Job_Requirement, and Required_Qualification columns respectively were dropped in separate copies of the dataset containing significant columns as they had no significant effect in individual copies created.

ANALYZING THE CLEANED DATASET

This is performed to give proper insight into the dataset in terms of its summary statistics and visualization. To solve the questions posed by our objectives, the following analysis were carried out on the dataset.

Technologies used include Google Colaboratory, Microsoft Power Bi, Microsoft Excel, and GitHub.

TEXT ANALYSIS USING NATURAL LANGUAGE PROCESSING (NLP) KITS

The dataset being mostly non-structural or textual had to be analyzed using NLP toolkits such as Textblob for processing the data to remove stopwords and retrieve noun phrases.

The Job_Description, Job_Requirement, and Required_Qualification columns were used processed to produce noun phrases that can be used as potential skills for analysis. Textblob was used to process these columns and extracted noun phrases in relation to top job titles and professions and years. The results obtained from processing the Required_Qualification column contained relevant noun phrases that were used as skills.

PROBLEM STATEMENT ANALYSIS

The Demand for Job Titles and Professions

An initial count to determine the year with the highest number of job postings revealed 2012 with the highest count of job postings. An average of 1544 jobs were posted every year.


Total Job Posted per Year

Similarly, March had the highest number of job postings with a steady decline in job postings from September to December. There were fewer job postings in January, April, and December compared with other months. An average of 1544 jobs are also posted every month.


Total Job Posted per Month
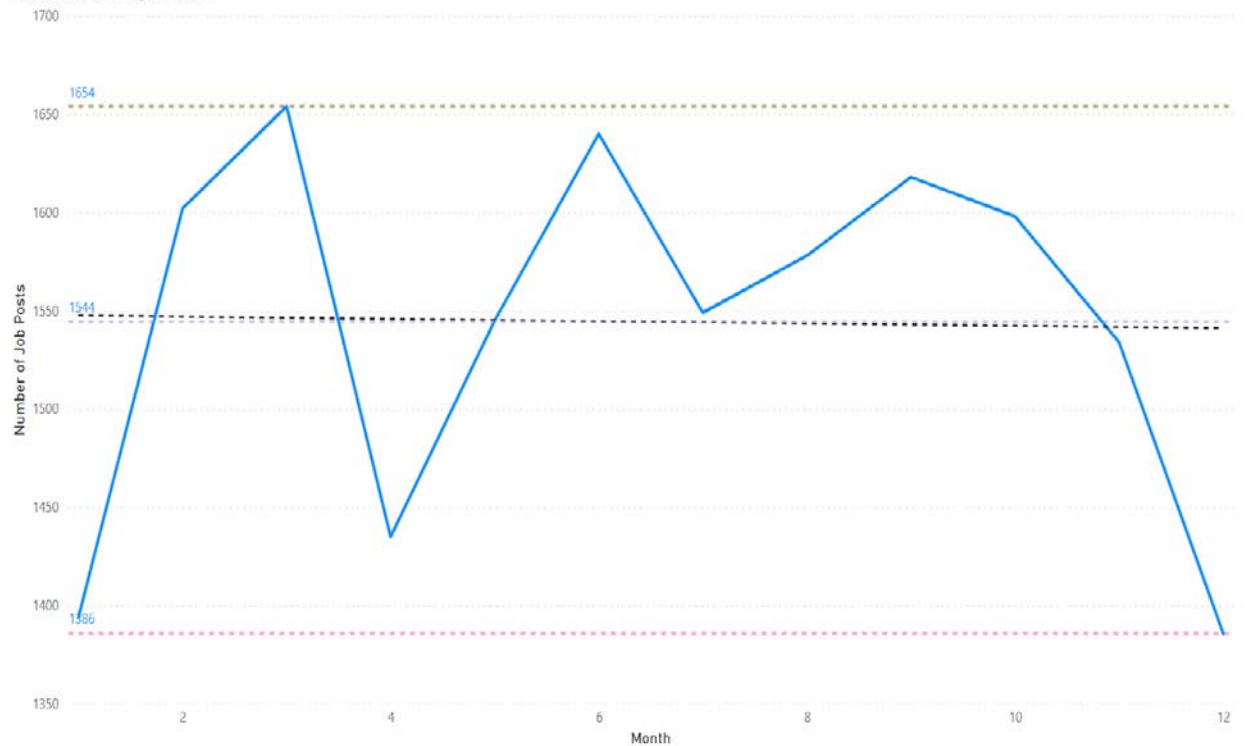
With the Title column, a count revealed the top most frequently posted jobs between 2004 and 2015. It was revealed that Accountant positions were in high demand over the years closely followed by Chief Accountant positions and Medical Representative positions, making up the top three jobs in demand.

Top Jobs Titles and Professions
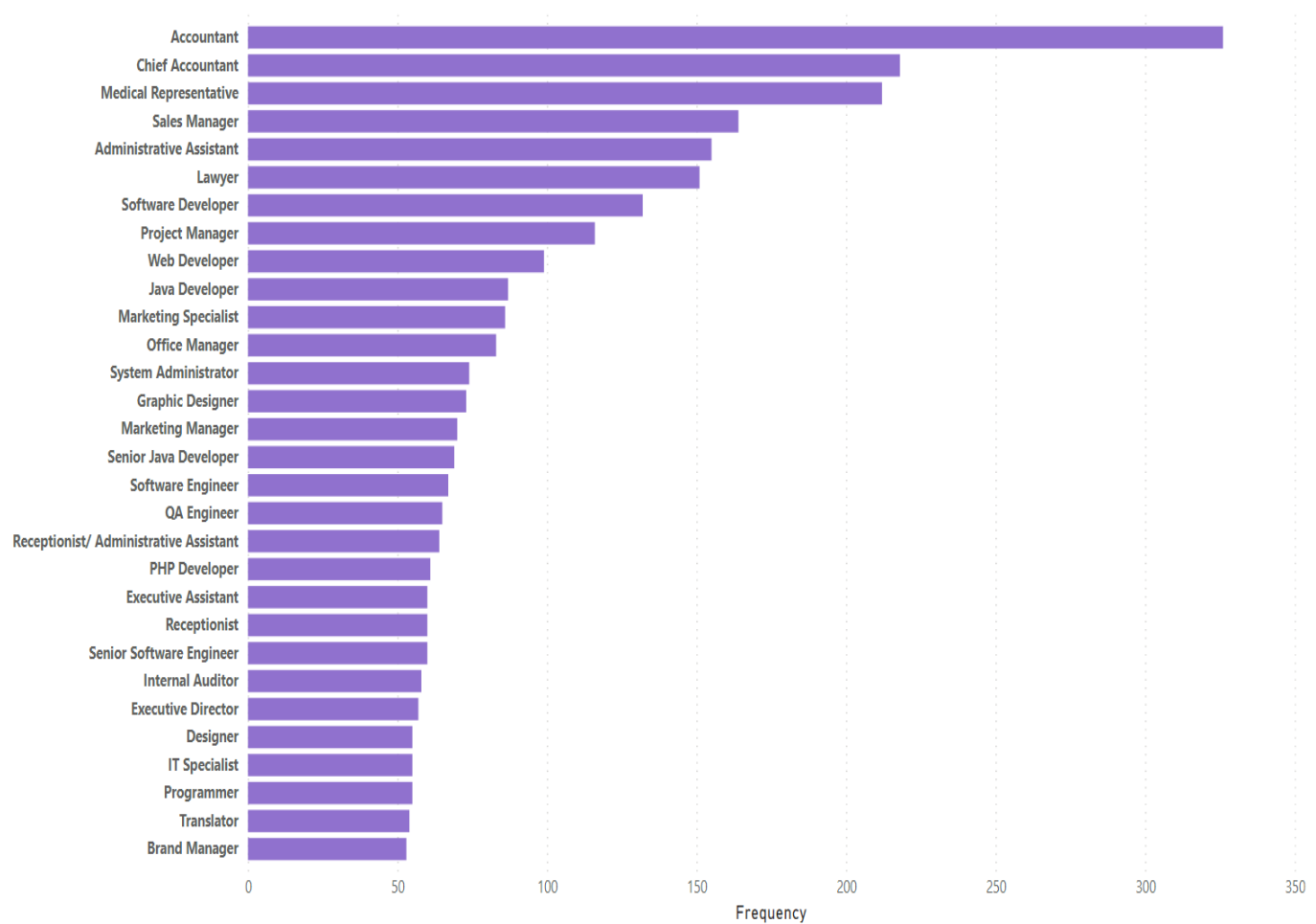
In terms of yearly job demand, Accountant positions were the highest in most years, followed by Medical Representative positions and Chief Accountant positions.

Most Demanded Jobs Titles or Professions per Year
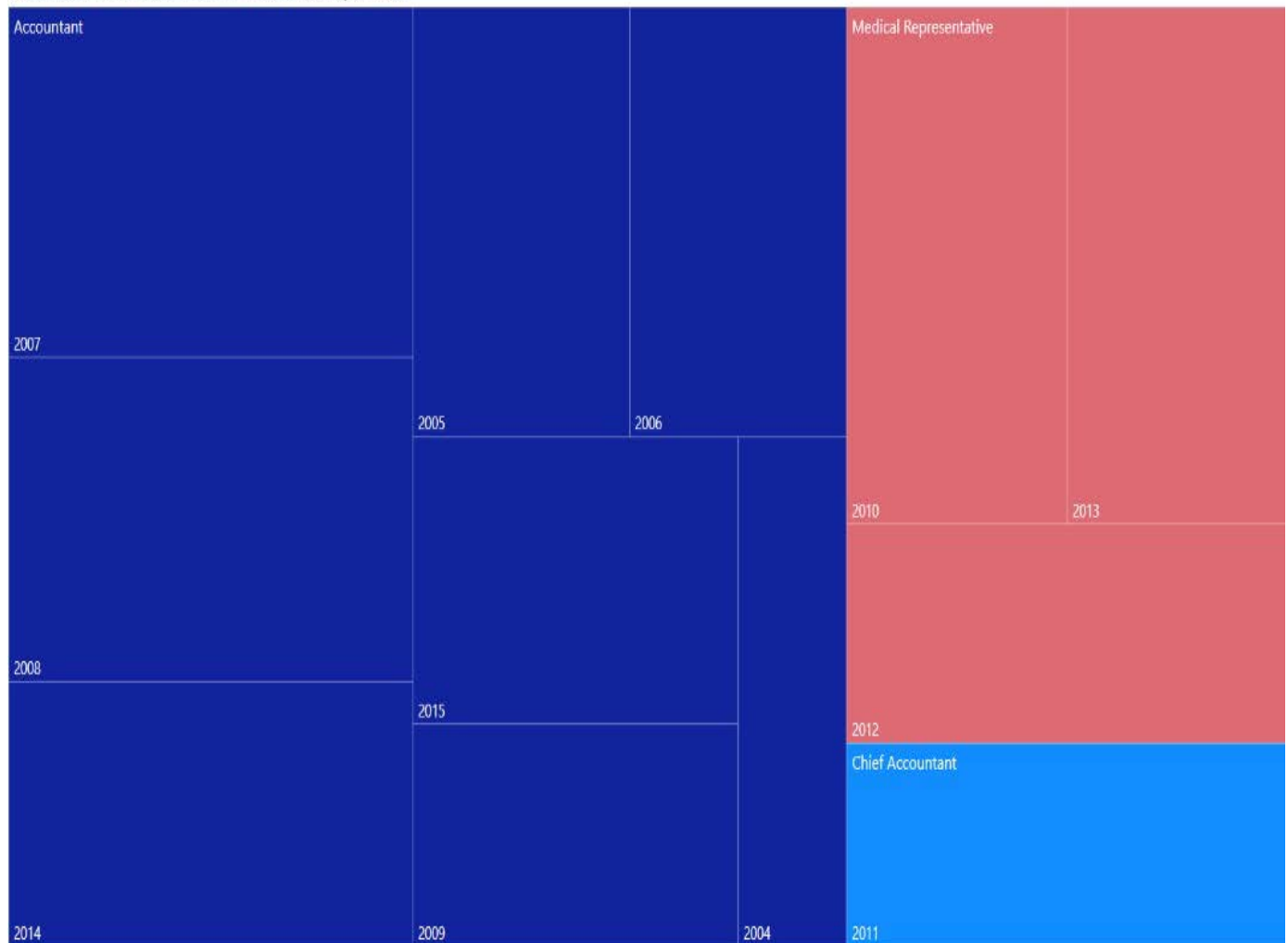
Furthermore, the demand for IT jobs was also observed to have been continuously lower than that of NON-IT jobs over the years, indicating a larger proportion of the populace in Armenia work in NON-IT industries.

Trend of Jobs Types
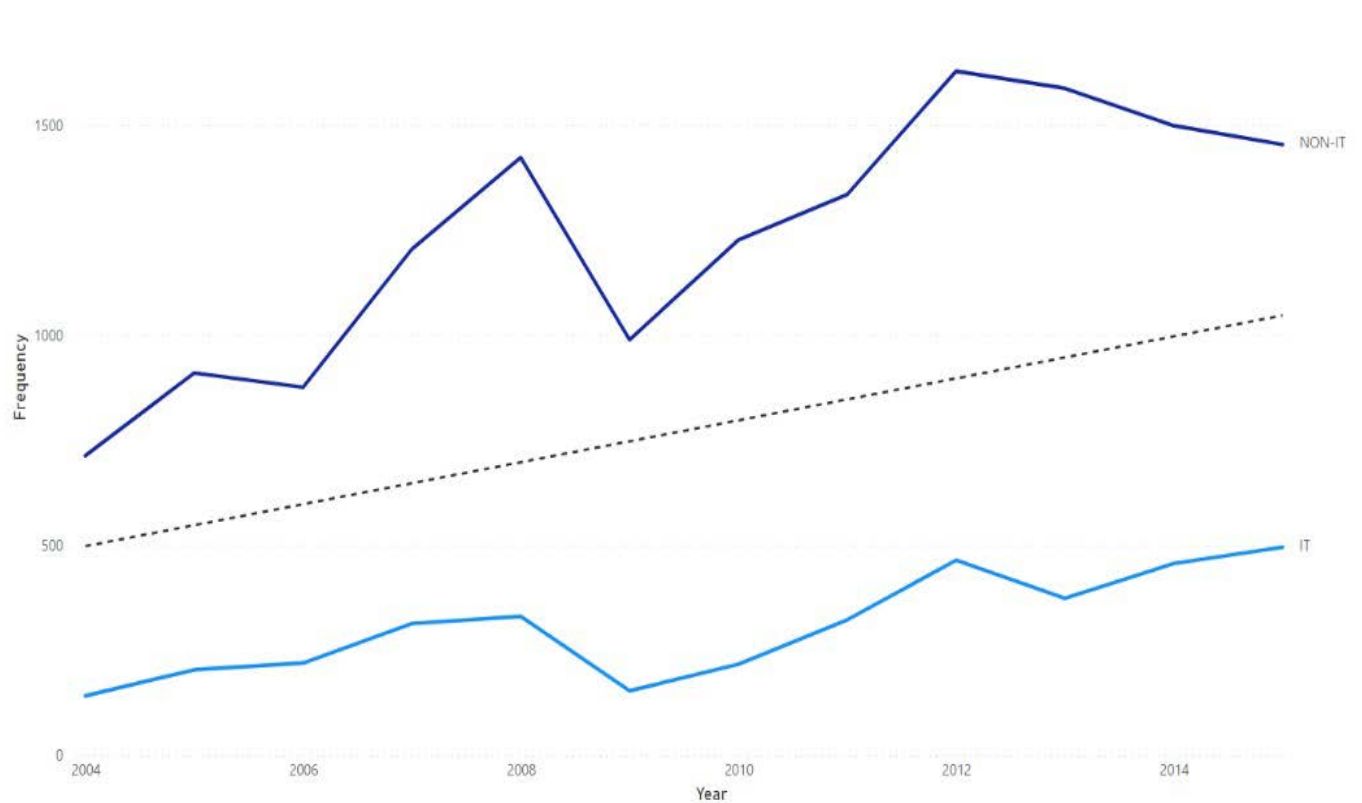
The Skills Required by Employers

English language and Microsoft Office Tools were the most in-demand skills in many of the jobs posted. Other skills that were required in the job posts include Russian languages, computer skills, analytical skills, communication skills, organizational skills. Microsoft Excel formed a large proportion of skills required among the Microsoft Tools.

Accountant positions also required Association of Chartered Certified Accountants (ACCA) certification.

Skills Required for an Accountant Position

Chief Accountant positions also required Association of Chartered Certified Accountants (ACCA) and a master's degree.

Skills Required for a Chief Accountant Position

Medical Representative positions also required language skills (English and Russian) more than Microsoft Office Tools. The position also required medical education and presentation skills.

Skills Required for a Medical Representative Position

The position of an Administrative Assistant required English language, Microsoft Office Tools, and communication skills.

Skills Required for an Administrative Assistant Position

A Lawyer position required English language, Microsoft Office Tools and communication skills, and a master's degree.

Skills Required for a Lawyer Position
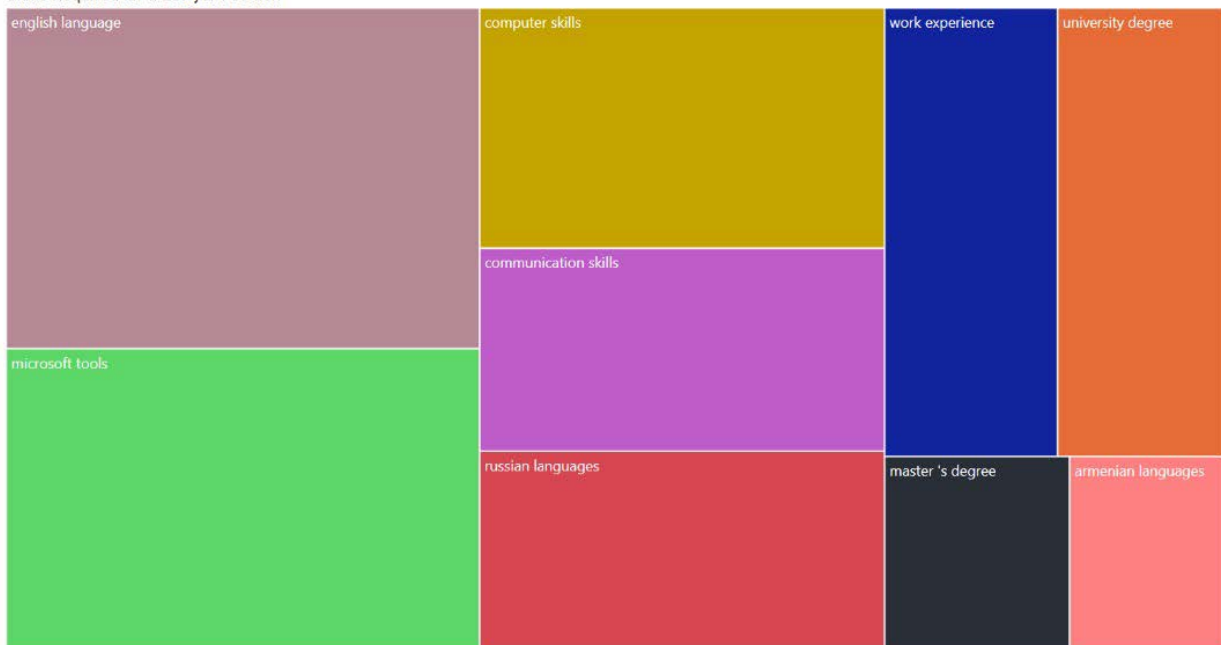
A Software Developer position required English language, and programming languages like
C#, SQL, C++, HTML, AJAX.

Skills Required for a Software Developer Position
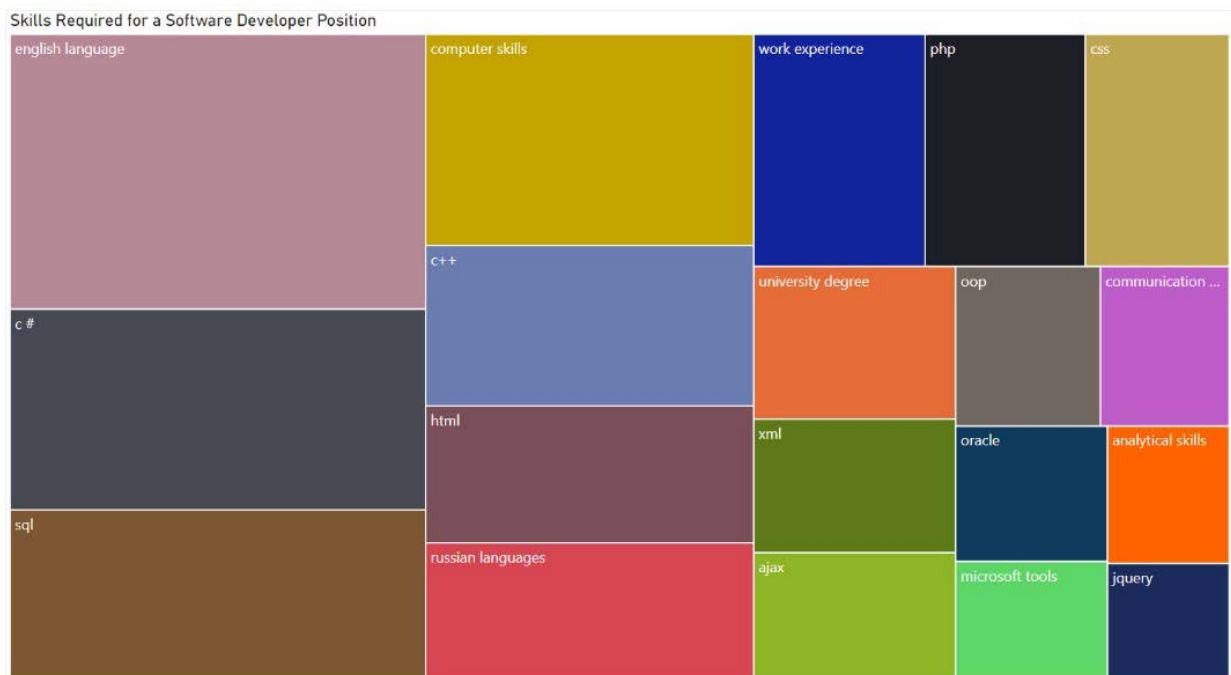
A Graphic Designer position required Adobe Illustrator, Abode Photoshop, and English
Language.

Skills Required for a Graphic Designer Position

The Trend of Necessary Skills by Year

Over the 12 years, the requirement for the English language was observed to have steadily risen as well as for Microsoft Office Tools and computer skills. The requirement for Russian languages was also in a steady rise until 2013, which was followed by a steady decline. The requirement for communication skills was also relevant all through the years.

Distribution of necessary skills over 12 years
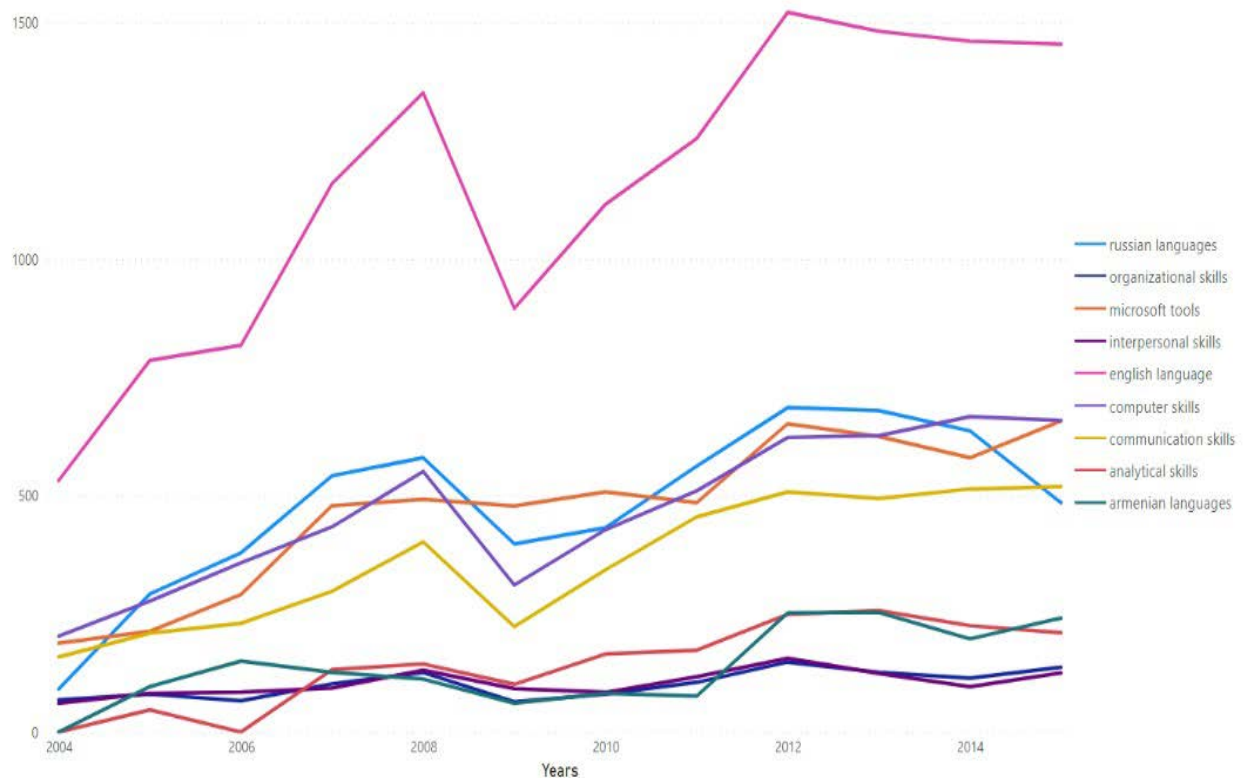
CONCLUSION

Based on the analysis, the month and year with most job postings were March and 2012 respectively. The month of April and September through to January recorded low job postings. Accountant positions were in high demand between 2004 to 2015.

Additionally, languages skills such as the English language, Russian languages, and Armenian languages are majorly required in many of the job postings. Microsoft Office Tools especially Microsoft Excel was also skills that were mostly required.

Lastly, English language, Microsoft Office Tools, computer skills, and communication skills remained relevant over the years, however, Russian languages were less required from 2013.

RECOMMENDATIONS

Inferences were made from the analysis performed and the following recommendations were proposed**:**

1. There are more opportunities in NON-IT related job positions as compared to IT positions, however Microsoft Office Tools especially Microsoft Excel and computer skills are required skills that are still relevant in both fields.

2. Languages especially the English language have always been relevant to every job. Fluency in the English language, as well as Russian languages and Armenian languages with improvement in communication skills, is highly recommended.

3. Job seekers looking to change careers should be more inclined towards the Finance related positions such as Accountant and Chief Accountant positions as they are highly demanded in the economy.

4. There is a decrease in job postings during April and September through to January, so the potential for stifler/tougher competition during this period could lead to difficulty in securing jobs. Targeted job applications in March have a higher chance of success.

5. Work experience was also seen to be required in many job posts. So, internships towards a particular job position are also advised.

6. Analysis with the 'Company' column can provide targeted insights into the demand for jobs and skills required by specific companies which can help in preparing a portfolio, resume, and application for the company.