Research Questions

1. What are the most common job titles and industries represented in job postings in New York City?

2. What are the most commonly required Technical skill and qualifications in job postings in New York City?

3. What soft skills are preferred?

Data

The data was sourced from City of New York's official jobs site.

It contains internal postings available to city employees and external postings available to the general public are included, as of February 24, 2023

There are 30 different columns categorizing each job according to the job title, job category, salary range, required skills and qualifications

Dataset contains 5,934 rows of data

Contains Job postings from 60 different agencies

Snapshot of the Dataset

Job ID	Agency	Posting Type	# Of Position	Business Titl	Civil Service	Title Classifi	Title Code N	Level	Job Category	Full-Time/Pa	Career Level	Salary Range	Salary Range	Salary Frequ	Work Locatic	Division/Wo	Job Descripti	Minimum Q	Preferred Ski	Add
533891	HRA/DEPT O	External	1	CONTRACT A	STAFF ANAL	Competitive-	12626	2	Administrati	F	Experienced	62470	76869	Annual	4 World Trac	Employment	The Departm	1. A master'	¢¬Ä¬ôs degre	**L
554785	DEPARTMEN	External	1	DESIGN ENG	CIVIL ENGINI		20202	0	Engineering,	F	Entry-Level	51413	62260	Annual	55 Water St	Pedestrian R	The Division	1. A baccalau	Experience r	As
568481	DEPT OF DES	Internal	1	Director	ADM PROC A	Competitive-	8297A	0	Finance, Acc	F	Experienced	65232	110599	Annual	30-30 Thoms	F&P/Acco/Ex	Only those w	1. A baccalau	Candidates shou	
527833	NYC HOUSIN	External	1	ELECTRICIAN	ELECTRICIAN		91717	0	Building Ope	F	Experienced	62.88	62.88	Hourly	Manhattan F	Manhattan F	Under the di	1. Five years of full-time s	1.F	
541555	FIRE DEPART	Internal	1	Director, Dat	CITY RESEAR	Non-Compet	21744	4A	Technology,	F	Experienced	94283	122290	Annual	9 Metrotech	Mgt. Info. Sy	The Fire Dep	1. For Assig	Bachelors or	***
558830	DEPT OF YOI	Internal	1	Program Ma	COMMUNITY	Non-Compet	56058	0	Constituent S	F	Experienced	54100	83981	Annual	2 Lafayette S	OEO: Advanc	The New Yor	1. A baccalau	Experience in an	
229124	DEPT OF HE/	External	1	HIV Program	SUPERVISIN(	Competitive-	51193	0	Health	F	Experienced	48900	69042	Annual	42-09 28th S	PHC Adminis	The Bureau c	1. A baccalau	Spanish speaking	
558052	DEPT OF EN\	External	1	Assistant Cor	AGENCY ATT	Non-Compet	30087	1	Legal Affairs	F	Experienced	63228	72712	Annual	59-17 Junctic	LEGAL	The NYC Dep	1. Admission	Excellent wri	As
558439	LAW DEPART	External	1	Paralegal - S	PARALEGAL .	Competitive-	30080	2	Legal Affairs	F	Experienced	43197	49677	Annual	100 Church S	Special Fede	The NYC Law	1. A baccalau	Additional Q	Due
558106	DEPT OF HE/	Internal	4	Public Health	PUBLIC HEAL	Competitive-	31215	1	Health Public	F	Experienced	45722	45722	Annual	NYC - All Bor	Food Safety	The Bureau c	1. A baccalau	The position	**L
558698	HRA/DEPT O	External	1	ADMINISTRA	PRINCIPAL A	Competitive-	10124	3	Social Servic	F	Experienced	58695	89699	Annual	4 World Trac	Office Of Cor	Under gener	1. A baccalau	¢¬Ä¬¢Know	The
553371	OFFICE OF T	Internal	1	Claims Speci	CLAIM SPECI	Competitive-	30726	3	Legal Affairs	F	Experienced	60132	69152	Annual	1 Centre St.,	Law and Adji	NOTE: This p	1. A baccalau	- Demonstrated	
548452	HRA/DEPT O	Internal	1	RECRUITER	PRINCIPAL A	Competitive-	10124	3	Administrati	F	Experienced	58695	67499	Annual	4 World Trac	Office Of HR	The Departm	1. A baccalau	¢¬Ä¬¢Abilit	**L
547026	DEPARTMEN	Internal	1	Chief Inform	IT PROJECT S	Non-Compet	95710	0	Technology,	F	Experienced	0	160000	Annual	33 Beaver St	Admin Execu	The NYC Dep	1. A baccalau	¢¬Ä¬¢Senic	APF
468504	DEPT OF HE/	Internal	1	Public Health	PUBLIC HEAL	Competitive-	31215	1	Health Public	F	Experienced	45722	52580	Annual	NYC - All Bor	Food Safety	Open for per	1. A baccalau	The position	**L
551133	DEPT OF EN\	Internal	1	Director of H	ADM MANAC	Competitive-	1002C	3	Administrati	F	Experienced	74730	130000	Annual	59-17 Junctic	EXECUTIVE A	PLEASE NOT	1. A baccalau	¢¬Ä¬¢Work	As
549840	ADMIN TRIA	External	2	Settlement C	EXECUTIVE A	Non-Compet	95005	M1	Legal Affairs	F	Experienced	110000	115000	Annual	66 John Stree	Hearing Divis	ABOUT OATI	Admission tc	1.Have significar	
544398	DEPARTMEN	External	1	Outreach Lia	COMMUNITY	Non-Compet	56058	0	Constituent S	F	Experienced	54100	83981	Annual	55 Water St	Office of Bic	The Division	1. A baccalau	Professional	***
532666	ADMIN FOR	Internal	1	Child Protect	DIRECTOR O	Non-Compet	95600	M1	Social Servic		Manager	58700	102226	Annual	1200 Waters	Bronx Zone C	The Protectiv	1. A baccalau	The preferre	Sec
540855	DEPT OF CIT	Internal	1	Project Mana	CITY RESEAR	Non-Compet	21744	2	Engineering,	F	Experienced	75504	86830	Annual	1 Centre St.,	Office of Ene	DCAS's missi	1. For Assig	The preferred ca	
564942	DEPARTMEN	Internal	1	Strategy and	ADMIN COM	Competitive-	1002F	0	Constituent S	F	Experienced	64749	134280	Annual	1 Liberty Pla	Executive	The Departm	1. A baccalau	¢¬Ä¬¢Excellent	
549840	ADMIN TRIA	Internal	2	Settlement C	EXECUTIVE A	Non-Compet	95005	M1	Legal Affairs	F	Experienced	110000	115000	Annual	66 John Stree	Hearing Divis	ABOUT OATI	Admission tc	1.Have significar	
527295	DEPARTMEN	External	7	RMAS Social	SOCIAL WOR	Non-Compet	52613	0	Social Servic	F	Entry-Level	51079	58741	Annual	75-20 Astori	Adult Offend	The New Yor	Qualification	¢¬Ä¬¢Exper	THI
535225	FIRE DEPART	Internal	1	WTC CASE M	CASE - MAN/	Non-Compet	50959	1	Health	F	Experienced	84744	84744	Annual	9 Metrotech	Medical	**THIS POSI	1. A valid Ne	An Associate dep	
546105	DEPT. OF HO	External	1	INTAKE WOF	COMMUNITY	Non-Compet	56057	0	Social Servic	F	Experienced	38333	63794	Annual	300 Skillman	Adult Service	The Division	Qualification	¢¬Ä¬¢Shou	LO/
531513	DEPARTMEN	Internal	1	Solution Arcl	SENIOR IT AI	Non-Compet	95711	0	Technology,	F	Experienced	100000	180000	Annual	55 Water St	IT & Telecom	Under gener	1. A baccalau	(1) TOGAF, F	The
537650	DEPARTMEN	Internal	1	Deputy Direc	AGENCY ATT	Non-Compet	30087	1	Constituent S	F	Experienced	63228	72712	Annual	75-20 Astori	Adult Offend	The New Yor	1. Admission	1.Experience	The
563811	MANHATTAI	External	1	Community /	COMMUNITY	Non-Compet	56057	0	Constituent S	F	Entry-Level	38333	52000	Annual	211 East 43r	Manhattan C	Manhattan C	Qualification	¢¬Ä¬¢Baccala	
572795	DEPT OF PAF	Internal	1	Data Scientis	CITY RESEAR	Competitive-	21744	3	Technology,	F	Experienced	95000	105000	Annual	Arsenal 830	Asst Comm f	The Data Tea	1. For Assig	1.Successful	NO
572795	DEPT OF PAF	External	1	Data Scientis	CITY RESEAR	Competitive-	21744	3	Technology,	F	Experienced	95000	105000	Annual	Arsenal 830	Asst Comm f	The Data Tea	1. For Assig	1.Successful	NO
570538	LAW DEPART	Internal	1	File Clerk for	CLERICAL AS	Competitive-	10251	3	Building Operations & Ma	Experienced	19.9179	22.9053	Hourly	100 Church S	Admin. Servi	Our Facilities	Qualification	Requiremen		
572204	DEPARTMEN	External	1	ENVIRONME	SUMMER CC	Non-Compet	10234	0	Engineering,	P	Student	17.5	17.5	Hourly	120 Broadwa	Environment	Internship Pr	As of June of Undergradua	Pro	
563157	DEPARTMEN	External	2	Outreach Spe	COMMUNITY	Non-Compet	56058	0	Constituent S	F	Experienced	54100	83981	Annual	59 Maiden Li	Safety Educa	The NYC Dep	1. A baccalau	Experience prese	
528462	DEPARTMEN	Internal	8	Recreation P	RECREATION	Competitive-	60430	0	Constituent S	F	Entry-Level	41660	47909	Annual	75-20 Astori	Adult Offend	The New Yor	1. A baccalaureate degree	¢¬Ä¬¢	The
565691	DEPT OF EN\	Internal	1	Assistant Civ	ASSISTANT C	Competitive-	20210	0	Engineering,	F	Entry-Level	57078	85646	Annual	465 Columbu	SUTTON PAF	The NYC Dep	1. A baccala	¢¬Ä¬¢ Profi	The
503612	NYC HOUSIN	External	4	PLASTERER	PLASTERER	Competitive-	92235	0	Building Ope	P	Experienced	44.82	44.82	Hourly	Bronx Floatir	Bronx Proper	Under super	Education and Experience	1.F	
432393	DEPARTMEN	External	5	EXTERMINAT	EXTERMINAT	Competitive-	90510	1	Public Safety	F	Experienced	34027	47967	Annual	75-20 Astori	Assets Mana	The New Yor	A valid certificate as a Cer	For	
556124	DEPT OF EN\	External	1	Business Ana	COMPUTER S	Competitive-	13632	1	Engineering,	F	Experienced	81951	102136	Annual	96-05 Horace	PROJECT MA	The NYC Dep	(1) A baccala	¢¬Ä¬¢1 ¢¬	***
551319	DEPARTMEN	Internal	1	Project Mana	CITY PLANNE	Competitive-	22122	2	Engineering,	F	Experienced	65208	95993	Annual	55 Water St	Franchise/Cc	The Office of	1. A baccalau	Excellent org	As
366589	ADMIN FOR	Internal	1	Agency Impl	PROGRAM E	Competitive-	52416	0	Social Servic	F	Experienced	61850	76745	Annual	150 William	Title IV-E	Under mana	1. A master's	The preferre	Sec

NYC_Jobs

Methodology

For preprocessing, text data was converted into lowercase. Numbers, special/junk characters, and NLTK defined stop words were deleted, and the text was lemmatized.

To identify the necessary skills from our data, we use **classification**.

- The data is clustered based on job title
- We opt for **multilabel classification** as there may be multiple job titles with similar characteristics.
- X**GBoost Classifier** was used to combine the outputs from individual decision trees to perform regression or classification.This model helps prevent overfitting by combining many decision trees.
- The data is split 80-20 for train/test data. For feature extraction, **CountVectorizer** is used in a format supported by machine learning algorithms from text datasets.

To figure out what the minimum requirements were for most jobs, we summed up the amount of listings a requirement appeared in a listing as the lowest accepted qualification

For finding the specific soft skills employers were seeking, we performed feature extraction on our classified data.

- Technical skills were extracted by matching predefined sets of technical skills to the words in each class.
- To extract soft skills, we leverage the **TextBlob library**, which uses POS-tags to identify adjectives that describe a person's skills or traits.
- The most significant characteristics of each class were determined by computing the probabilities of each feature within that class, and then mapping the tokens to their corresponding probabilities.
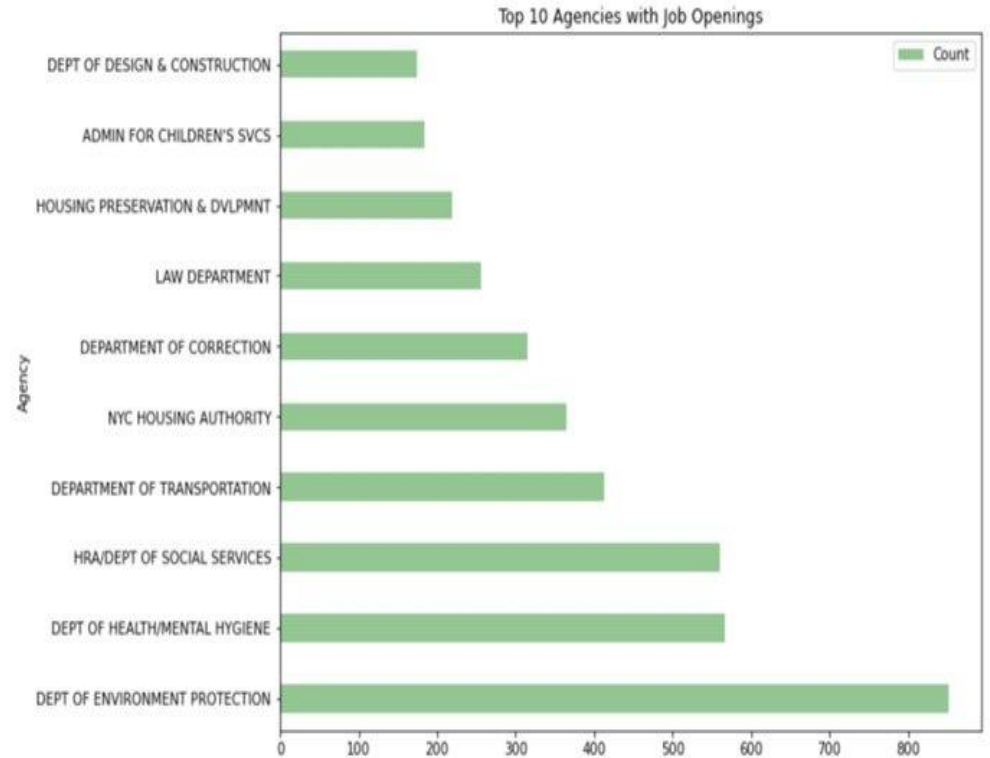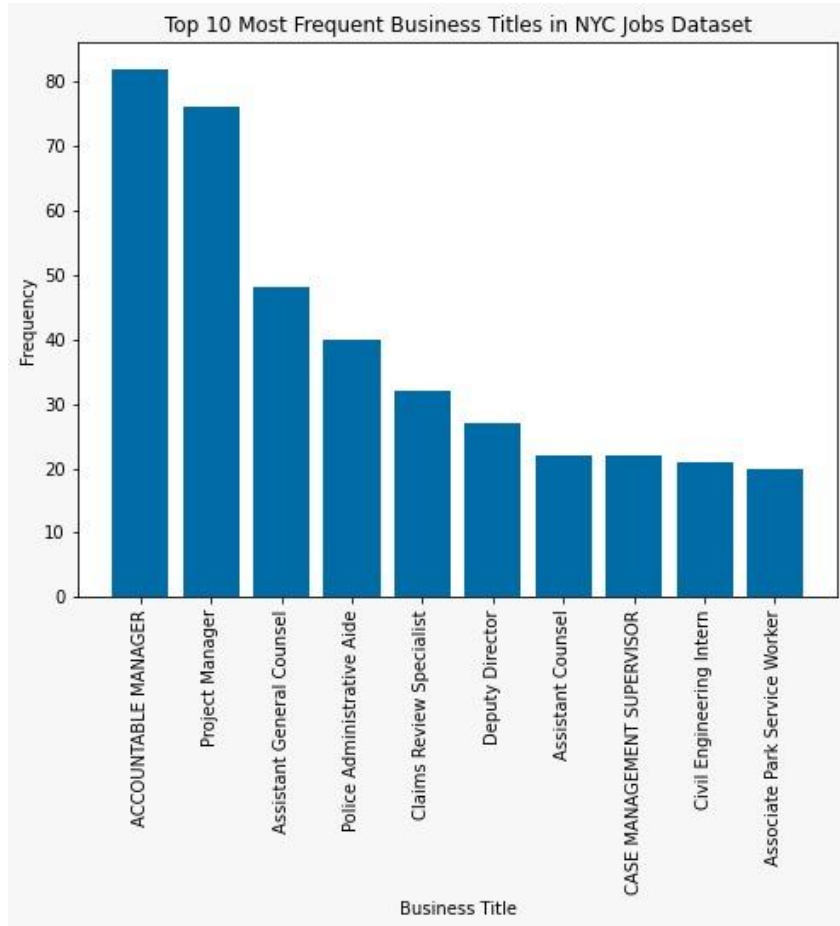
Classification Results

```
In [259]:  1  #evaluate the predictions
           2  print("Accuracy score is: ",accuracy_score(y_test, y_predicted))
           3  print("Classes: (to help read Confusion Matrix)\n", clf.classes_)
           4  print("Confusion Matrix: ")
           5
           6  print(confusion_matrix(y_test, y_predicted))
           7  print("Classification Report: ")
           8  print(classification_report(y_test, y_predicted))
```

```
Accuracy score is:  0.8171861836562764
Classes: (to help read Confusion Matrix)
 ['.NET Developer'
 '.NET Developer, Bureau of Environmental Disease and Injury Prevention'
 '.Net Programmer/Analyst for the Division of HPD Tech' ...
 'ZONING DIVISION SUMMER COLLEGE INTERN' 'ZONING SPECIALIST'
 'ZONING URBAN DESIGNER']
Confusion Matrix:
[[4 0 0 ... 0 0 0]
 [0 1 0 ... 0 0 0]
 [0 0 1 ... 0 0 0]
 ...
 [0 0 0 ... 1 0 0]
 [0 0 0 ... 0 1 0]
 [0 0 0 ... 0 0 1]]
Classification Report:

precision    recall  f1-score   support
```

Classification Results (Cont)

Top 10 Most Frequent Business Titles in NYC Jobs Dataset

Top 10 Agencies with Job Openings

Feature extraction

```
In [478]:  1  print(output.T)
```

```
                                                                       0  \
job_title                                                  .NET Developer
soft_skills           [net, full, strategic, limited, required, visu...
technical_skills                                      [excel, software]

                                                                       1  \
job_title             .NET Developer, Bureau of Environmental Diseas...
soft_skills           [healthy, technical, provide, initial, develop...
technical_skills                                [excel, software, data]

                                                                       2  \
job_title             .Net Programmer/Analyst for the Division of HP...
soft_skills           [programmer, web, programmer, affordable, larg...
technical_skills                        [excel, programming, ui, react]

                                                                       3  \
job_title             ASSISTANT DEPUTY COMMISSIONER, INFRASTRUCTURE ...
soft_skills           [technical, individual, direct, dhs, administr...
technical_skills                     [cloud, analysis, software, excel]

                                                                       4  \
job_title             ASSISTANT DEPUTY COMMISSIONER, NETWORK,TELEPHO...
soft_skills           [technical, direct, administrative, commission...
technical_skills                         [excel, analysis, software]

                                                                       5  \
job_title                                             AUTOMATION INTERN
soft_skills           [financial, write, functional, financial, unde...
technical_skills      [excel, software, database, java, analysis, ui...
```
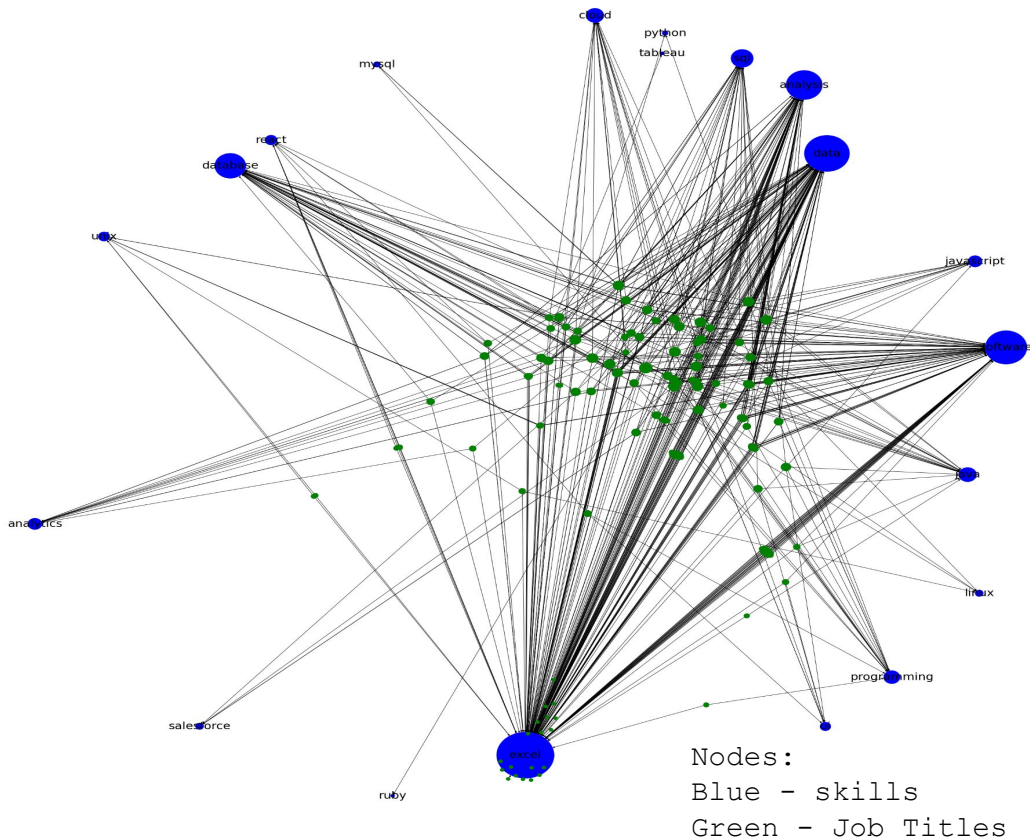
Minimum Requirement Findings:

Out of the 5934 listings, the amount of times a degree was the minimum requirement or was an alternative requirement:
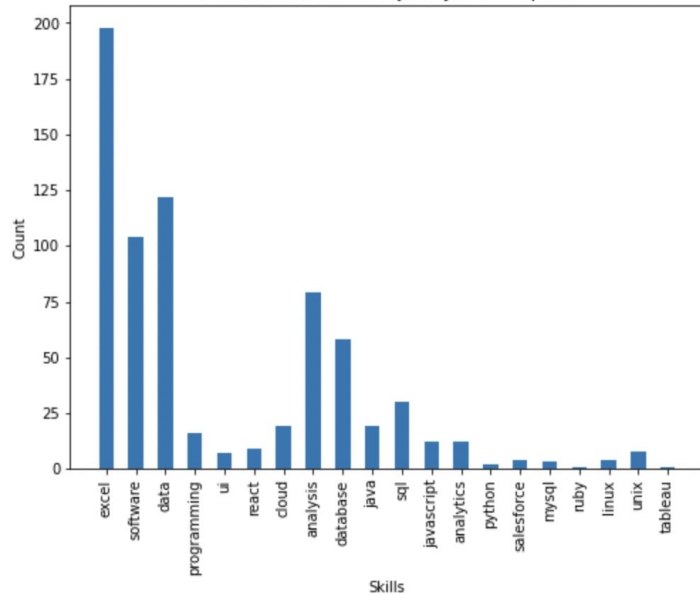
- **High school education** was mentioned in 2661 listings (45% of listings), out of which 99% require some form of experience. 634 listings ONLY mentioned high school for educational requirements.
- 241 listings had **Associates Degree** as the minimum, and ALL listings require some form of experience. 912 listings mentioned the degree.
- 3610 listings had a **Bachelor's Degree** as the minimum (61% of listings), and 38% of them require some form of experience. 3622 listings mentioned the degree.
- 531 listings had a **Master's Degree** as the minimum, and 96% of them require some form of experience. 1713 listings mentioned the degree.
- 32 listings had a **Doctorate** as the minimum, and none of them required any additional experience. 306 listings mentioned the degree
- Most common **minimum additional experience** required was one year (55% of listings), followed by six and five years (10.6%/8.5%).
- Listings allowed applicants to substitute experience with either their time within an educational system or a certain amount of credits counting as a year, varying between listing.

Technical skills from Tech jobs (Job Description)

Nodes:
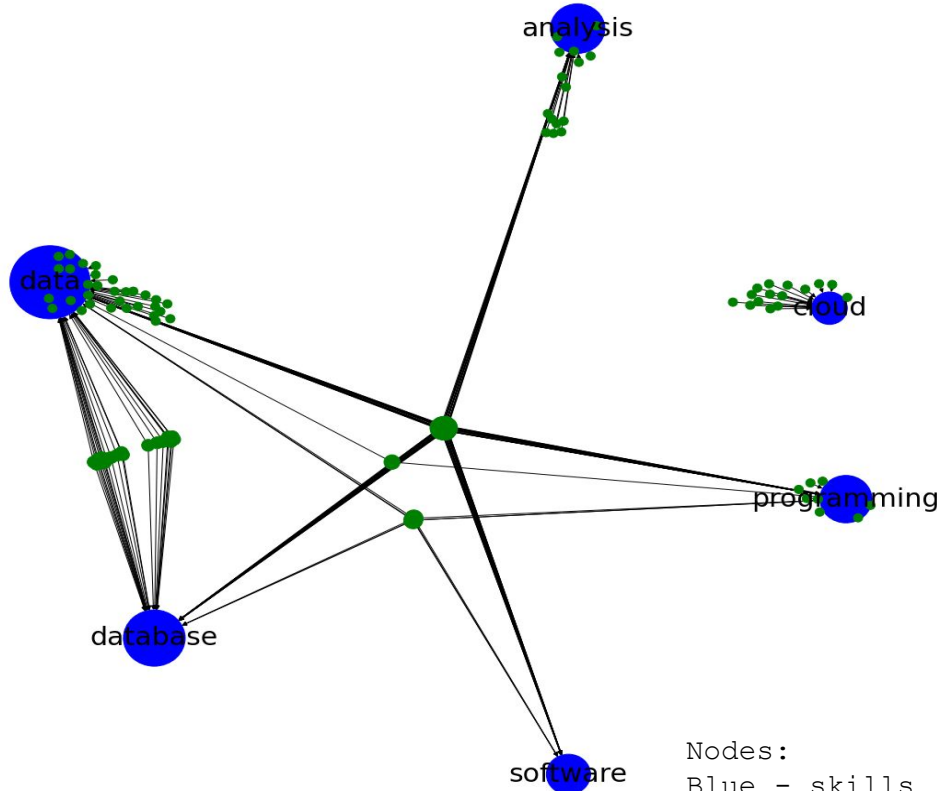Blue - skills
Green - Job Titles

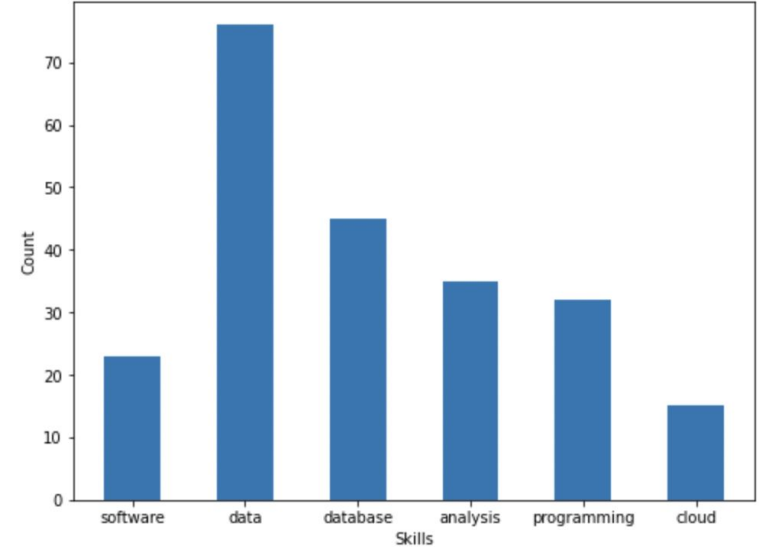Excel, data analysis and software programming are the high demand skills

Technical skills from Tech jobs (Minimum Qualifications)

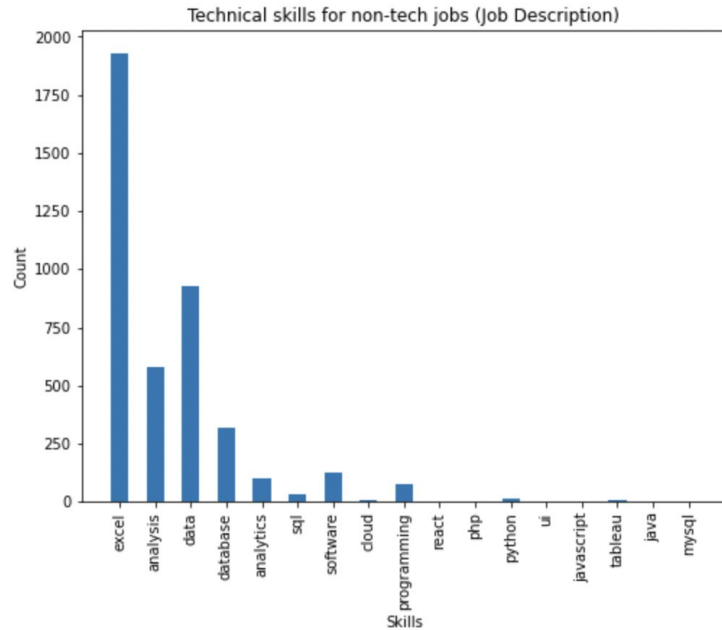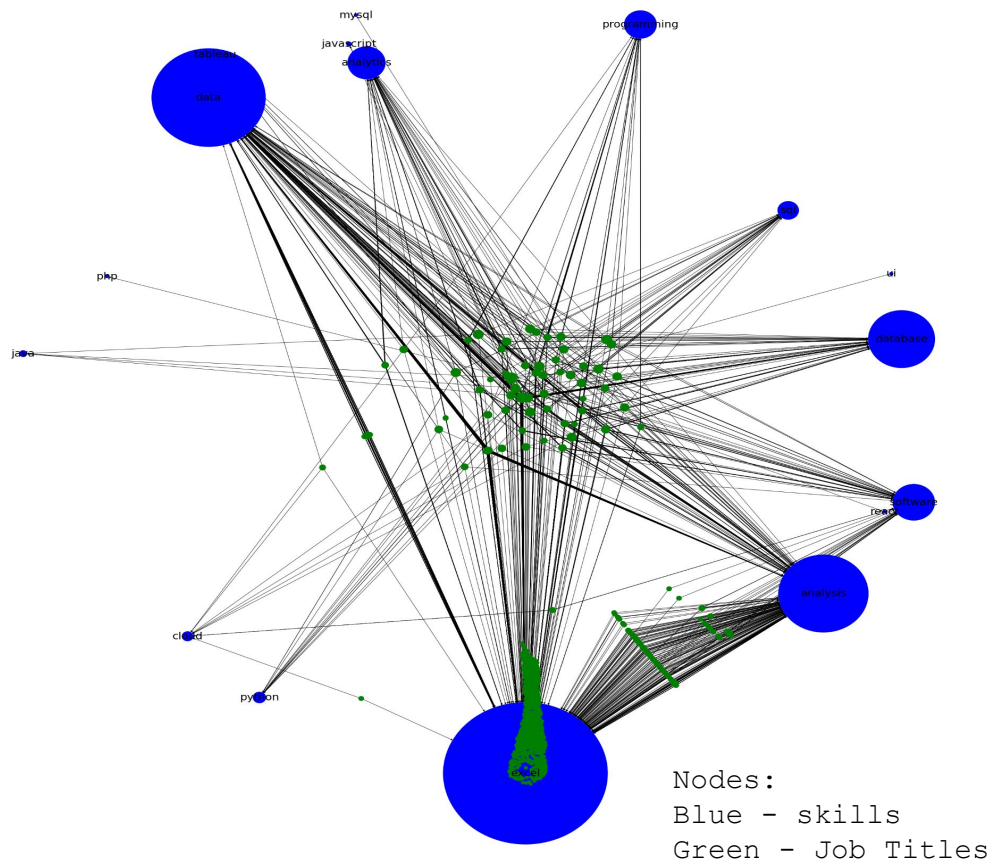Technical skills for tech jobs (Minimum Qualifications)

Minimum required skills for tech jobs are data analysis, software programming, database and cloud
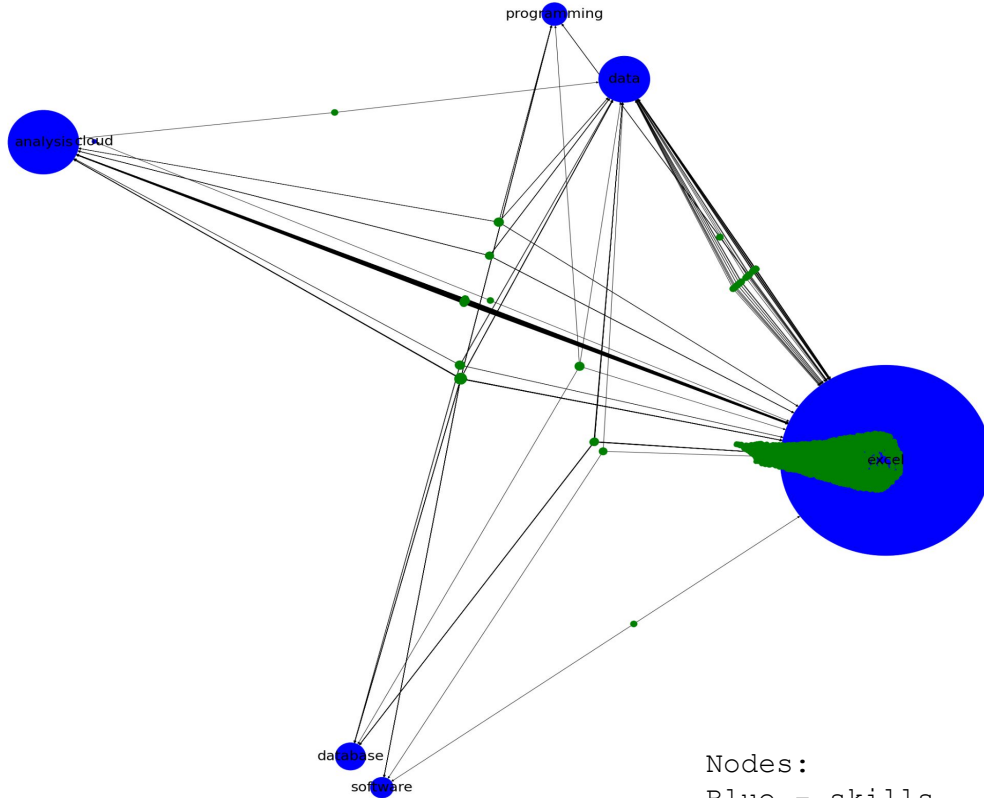
Nodes:
Blue - skills
Green - Job Titles

Technical skills from Non-Tech jobs (Job Description)
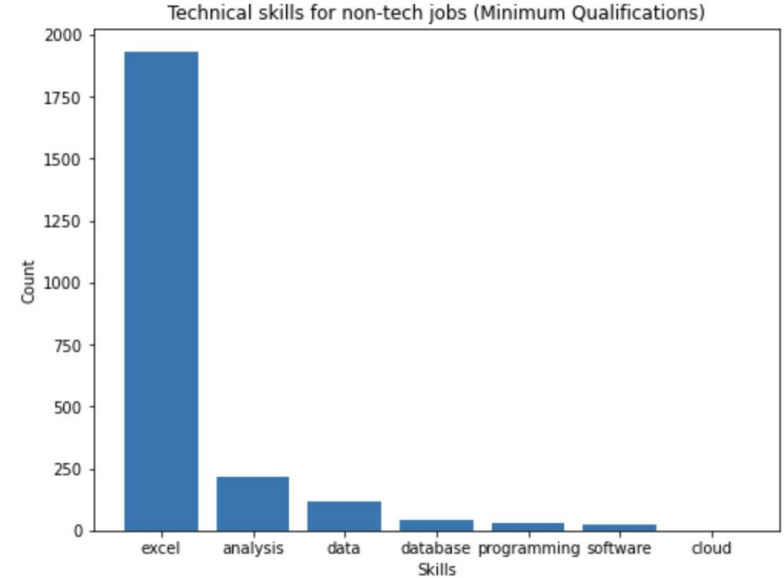

Nodes:
Blue - skills
Green - Job Titles

Excel has the highest demand
followed by data analysis

Technical skills from Non-Tech jobs (Minimum

Technical skills for non-tech jobs (Minimum Qualifications)

Nodes:
Blue - skills
Green - Job Titles

Excel is the minimum requirement
for most of the Non technical jobs

Soft Skill Extraction Findings (Tech):

Bold is overlap with non-tech, italics is a specific trait. Some traits are clustered differently/multiple times.

- Communication: **responsible**, consultative, *social*, individual, supervisory, substitute, **inquiry**, answer, centered
- Adaptability: wide, updating, *specialized*, relevant, recent, eligible, enhancement, appropriate, moderate
- Teamwork: *trained*, top, *organized*, enterprise, appointed
- Problem-solving: valid, *technical*, *statistical*, *scientific*, specific, pertinent, relevant, *fiscal*, condition, console, cellular
- Time management: twenty, minimum, **maximum**, semester, minute
- Leadership: *supervisory*, top, **responsible**, executive, appointed
- Creativity: visual, ***innovative***, enhancement
- Attention to detail: specific, superficial, statutory, relevant, appropriate, moderate, legal, minimum, **maximum**
- Emotional intelligence: *social*, individual, supervisory, ***responsible***, appropriate, moderate, eligible, appointment, organizational
- Conflict resolution: *social*, supervisory, ***responsible***, appropriate, moderate, eligible, appointment, organizational
- Misc: *cultural awareness*, goal setting, *self-confidence*, *work ethic*, *self-awareness*

Soft Skill Extraction Findings (Non-Tech):

Bold is overlap with tech, italics is a specific trait. Some traits are clustered differently/multiple times.

- Communication skills: meaningful, manner, liaise, lingual, interview, *interpersonal*, initiative, **inquiry**, *insight*, interactive, *communicate*.
- Leadership skills: manage, maintain, leading, *lead*, largest, larger, leading, *involve*, inspiration, *initiative*.
- Time management skills: mandatory, mandated, maintain, late, last, workload, *timely*, frequent, full.
- Adaptability skills: *flexible*, *adaptable*, improvise.
- Problem-solving skills: *logical*, logic, analyze, investigative, insight, insect, **inquiry**, ***innovative***, implementation, *impartial*, *impactful*.
- Teamwork skills: *collaborative*, teamwork, *coordination*, involve, *inclusive*, incorporated, interact, *interdependent*, involve.
- Creativity skills: ***innovative***, inspiration, *motivated*.
- Attention to detail skills: measurement, **maximum**, manual, little, list, limit, license, *logical*, layer, lowest, lower, lockbox, lock, lease.
- Emotional intelligence skills: *empathy*, *equitable*, *inclusiveness*, *integrity*, *inspire*, *inclusionary*, initiative.
- Work ethic skills: ***responsible***, *reliable*, *proactive*, *productive*, *professional*, *dependable*, *respectful*, fulfillment, *functional*. Individual trait for tech.

Conclusion

- Bachelor's Degrees are the most commonly asked for credential and are the least likely to require additional experience besides Doctorates.
- A high school education can provide a job if you have had substantial experience within the field.
- 'Excel' is required for both tech and non tech jobs.
- Data Analytics skill is in high demand.
- Soft skills in non-tech jobs are focused on individual traits and social aptitude with others, while for tech jobs the focus is on mechanical skill/knowledge with the term "social" being the umbrella term used for social and most personal traits. Both prioritize responsible and innovative employees.
- The method can be used on wide range of text data, in the form of job title and description.
- For international students, in the data there's no specification for visa sponsorship.
- Future work:
  Using bigrams and trigrams to eliminate error.