



JOB DATA ANALYSIS FOR SOME TECH JOBS



AIM OF PROJECT



The problem to be solved is to provide valuable insights into the job market, and assist employers, educational tech providers and recruiters in making informed decisions.



The goal is to carry out an Exploratory Data Analysis and develop a machine learning analysis system that collects job related information from various UK job websites.



BENEFITS OF THE ANALYSIS AND MODEL



Job seekers are able to carry out efficient job search



Educational tech providers will be able to understand the skills that are in high demand, hence having a guide in developing relevant programmes to address the gap in the job market.



To provide employers and recruiters with valuable data-driven insights that will enable them make informed decision, remain competitive and improve recruitment process.



DATA SOURCING



The dataset for this analysis was scrapped from various UK jobsites by the team of Data engineers, it was retrieved from the PostgreSQL server for analysis.

The initial data set provided had about 32,000 rows, which was later replaced with another set of data, because of the constraints which included, inconsistencies in data set, lots of missing data especially in the salary column, also the data containing specified tech jobs was too small for analysis.

The Data Engineers looked into it and based on our recommendation, provided new sets of data containing about 40,868 rows for analysis.



STEPS OF ANALYSIS





FEATURES OF THE DATASET



Job Title -The title of the job position

Job City -City where the job is located

Date posted -The date when the job was posted

Company name -The name of the company or organisation offering the job

Job type -The type of job whether full time, part time, contract etc

Salary -Salary range for the listed job

Job link -job link

Job description -Description of the job title

The tech jobs titles were group into 13 categories, Data Engineer, Data Scientist, Data Analyst, Power Platform, IT Infrastructure and Cloud Computing, Software Engineer, Business Analyst, Product Manager, Project Manager, CyberSecurity, Gaming and Multimedia, UI/UX Designer and Software Tester.



DATA OVERVIEW

```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 40869 entries, 0 to 40868
Data columns (total 10 columns):
#   Column          Non-Null Count  Dtype
---  -
0   job_title       40869 non-null  object
1   job_description 39911 non-null  object
2   job_link        40868 non-null  object
3   job_city        40714 non-null  object
4   job_country     36631 non-null  object
5   date_posted    24612 non-null  object
6   salary         24853 non-null  object
7   company_name    40264 non-null  object
8   job_type       20039 non-null  object
9   company        20 non-null     object
dtypes: object(10)
memory usage: 3.1+ MB
```

Insights: The data set has 40,869 entries, and data types all appear as object.

```
df.describe(include = 'all').T
```

	count	unique	top	freq
job_title	40869	13434	Business Analyst	717
job_description	39911	19888	About Us We're an award-winning innovative tec...	344
job_link	40868	40868	https://uk.linkedin.com/jobs/view/game-develop...	1
job_city	40714	4609	London	10322
job_country	36631	4	United Kingdom	21578
date_posted	24612	13842	2023-08-03	1408
salary	24853	7826	Competitive	2954
company_name	40264	6080	Michael Page	372
job_type	20039	47	Permanent	8080
company	20	18	UNDERSTANDING RECRUITMENT	2

Insights: The general overview shows that **Business Analyst** is most advertised tech career, with **London** as the city with the highest job vacancy

Data set has 40,868 rows and 10 columns, with lots of missing data.



DATA PREPROCESSING

A copy of the original data was made. The Job link column was dropped, since its a unique value. Duplicates were found and dropped. There were some missing data in the dataset, the company column has almost 100% missing data, it was dropped. The job type has about 67% missing data (above 50%), this will be filled with the mode, but may affect the integrity of the data. The rest of the columns have missing data below 22% and were filled with mode values.

Out[16]:

	job_title	job_description	job_city	job_country	date_posted	salary	company_name	job_type	company
25906	BI Analyst	BI Data & Analytics, Data Analysis & Insights\...	Leeds	United Kingdom	2023-08-04	Up to £35,000	Oscar Technology	Permanent	NaN
25964	BI Analyst	BI Data & Analytics, Data Analysis & Insights\...	Leeds	United Kingdom	2023-08-04	Up to £35,000	Oscar Technology	Permanent	NaN
25988	BI Analyst	BI Data & Analytics, Data Analysis & Insights\...	Leeds	United Kingdom	2023-08-04	Up to £35,000	Oscar Technology	Permanent	NaN
26004	BI Analyst	BI Data & Analytics, Data Analysis & Insights\...	Leeds	United Kingdom	2023-08-04	Up to £35,000	Oscar Technology	Permanent	NaN
26026	BI Analyst	BI Data & Analytics, Data Analysis & Insights\...	Leeds	United Kingdom	2023-08-04	Up to £35,000	Oscar Technology	Permanent	NaN
26046	BI Analyst	BI Data & Analytics, Data Analysis & Insights\...	Leeds	United Kingdom	2023-08-04	Up to £35,000	Oscar Technology	Permanent	NaN
26066	BI Analyst	BI Data & Analytics, Data Analysis & Insights\...	Leeds	United Kingdom	2023-08-04	Up to £35,000	Oscar Technology	Permanent	NaN

Insights: More duplicated rows noted above.



DATA PREPROCESSING

Job title was grouped into Job Categories, as seen in the diagram and new column called `job_category` was created

```
keywords = ['manager', 'tester', 'software tester', 'QA engineer, software tester', 'business', 'test automation', 'games', 'product']

def selected_jobs(job_title, keywords):
    job_title_lower = job_title.lower()

    if any(keyword.lower() in job_title_lower for keyword in keywords):

        if any(keyword.lower() in job_title_lower for keyword in ['cloud', 'devops', 'cloud engineer', 'devops engineer', 'archite
            return 'IT Infrastructure and Cloud Computing'
        elif any(keyword.lower() in job_title_lower for keyword in ['tester', 'software tester', 'QA engineer, software tester',
            return 'Software Tester'
        elif any(keyword.lower() in job_title_lower for keyword in ['power platform', 'power apps', 'power automate', 'power', 'p
            return 'Power Platform'
        elif any(keyword.lower() in job_title_lower for keyword in ['network security', 'cyber', 'security', 'network engineer', '
            return 'Cyber Security'
        elif any(keyword.lower() in job_title_lower for keyword in ['software', 'software engineer', 'programmer', 'java', 'javasc
            return 'Software Engineer'
        elif any(keyword.lower() in job_title_lower for keyword in ['game designer', 'game', 'games', '3D modeling', 'game program
            return 'Gaming and Multimedia'
        elif any(keyword.lower() in job_title_lower for keyword in ['business intelligence', 'bi', 'business analyst', 'business'])
            return 'Business Analyst'
        elif any(keyword.lower() in job_title_lower for keyword in ['machine learning', 'ai', 'artificial intelligence', 'python'
            return 'Data Scientist'
        elif any(keyword.lower() in job_title_lower for keyword in ['data engineer', 'data infrastructure', 'etl processes', 'dat
            return 'Data Engineer'
        elif any(keyword.lower() in job_title_lower for keyword in ['data', 'analyst', 'analytics']):
            return 'Data Analyst'
        elif any(keyword.lower() in job_title_lower for keyword in ['project', 'project management']):
            return 'Project Manager'
        elif any(keyword.lower() in job_title_lower for keyword in ['product', 'lead', 'product planning', 'product management'])
            return 'Product Manager'
        elif any(keyword.lower() in job_title_lower for keyword in ['design', 'designer', 'graphics', 'ui', 'ux']):
            return 'UI/UX Designer'
    else:
```



DATA PREPROCESSING

Job type was grouped into paid staff, volunteer and intern, under employment_type .

Rows without numeric values, in the salary column were located and dropped, bringing the number of rows to 12,777. To further clean the salary column, the pounds sign, comma was eliminated while the 'k' meaning thousand was replaced with '000'. Salary was split into three columns the minimum, maximum and the column containing duration. The unwanted text and delimiters were eliminated by extracting the numerical values from the min and max salary column . Another split function was written to split the column where the min and max salary still existed on the same column, and min_salary_new column was created, this was used as the salary column for analysis. All unwanted numbers including zeros seen after the space in the min_salary_new column was eliminated. The data type for the min_salary_new was changed to numerical.

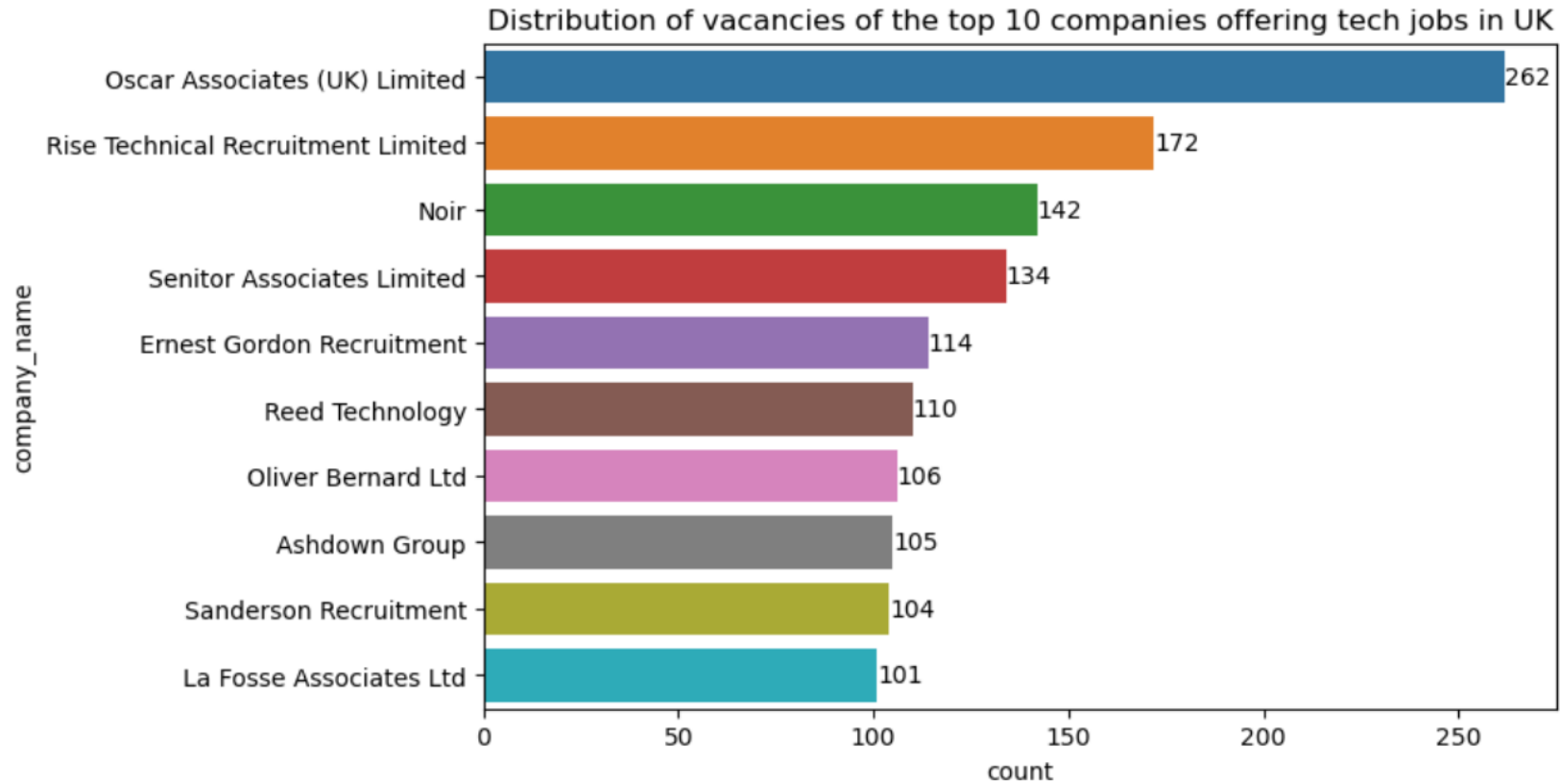
Salary duration column was created from the duration column, it shows the duration of the salary, whether it is per hour, day, month or year. Based on the salary duration , the yearly salary was calculated.

The different career level of the staff, was grouped and classified into entry-level, mid-level, senior-level, and managerial level position based on their yearly salary, Based on ai's suggestions, the skills for each job_category was identified and grouped as job_category_skills.





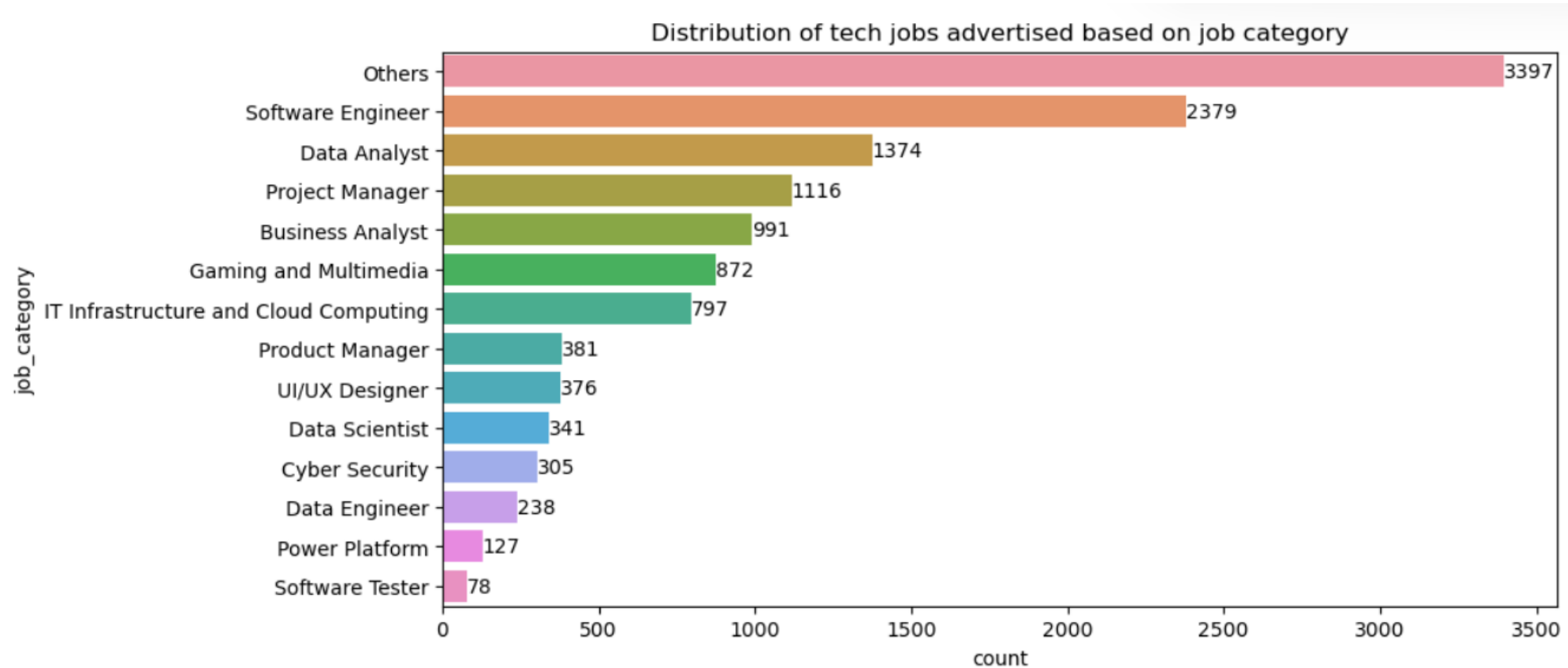
EXPLORATORY DATA ANALYSIS



Insights: Among the top ten companies with tech job vacancies, Oscar Associates (UK) Limited tops the chart, 2% (262) of the jobs advertised comes from them. Other companies recorded among the top ten includes Rise Technical Recruitment Limited and Noir.



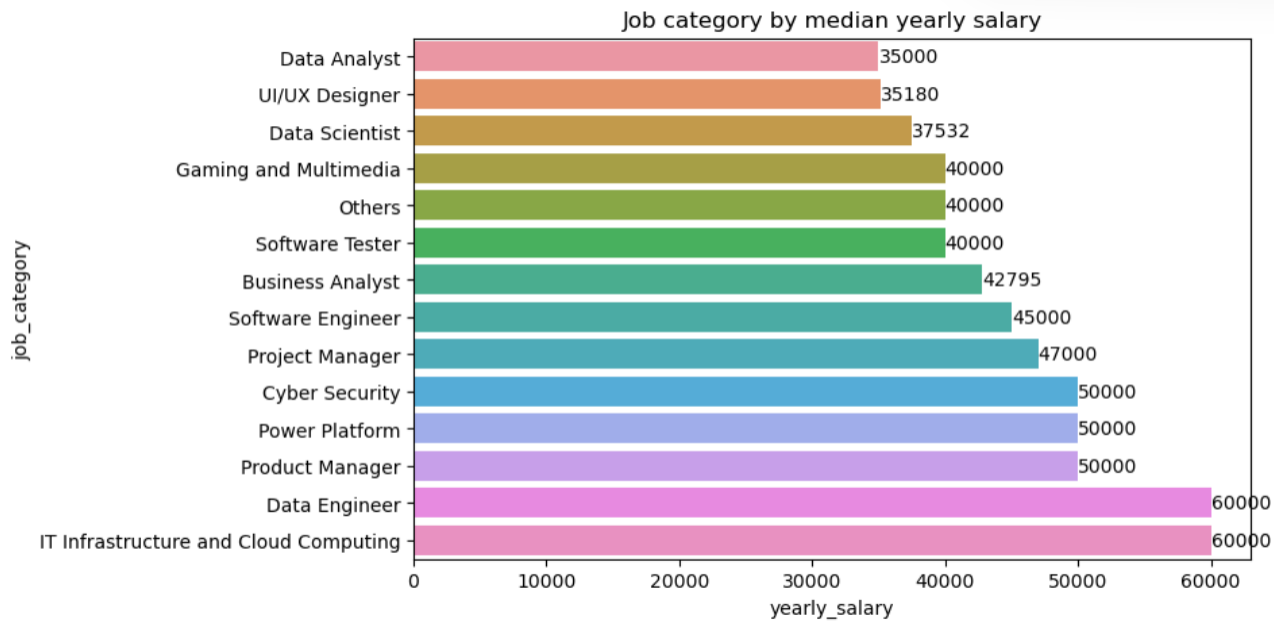
EXPLORATORY DATA ANALYSIS



Insights: Software Engineer, Data Analyst, Project Manager and Business Analyst are the top 4 tech jobs being advertised in United Kingdom, while Software Tester and Power Platform are the least



EXPLORATORY DATA ANALYSIS

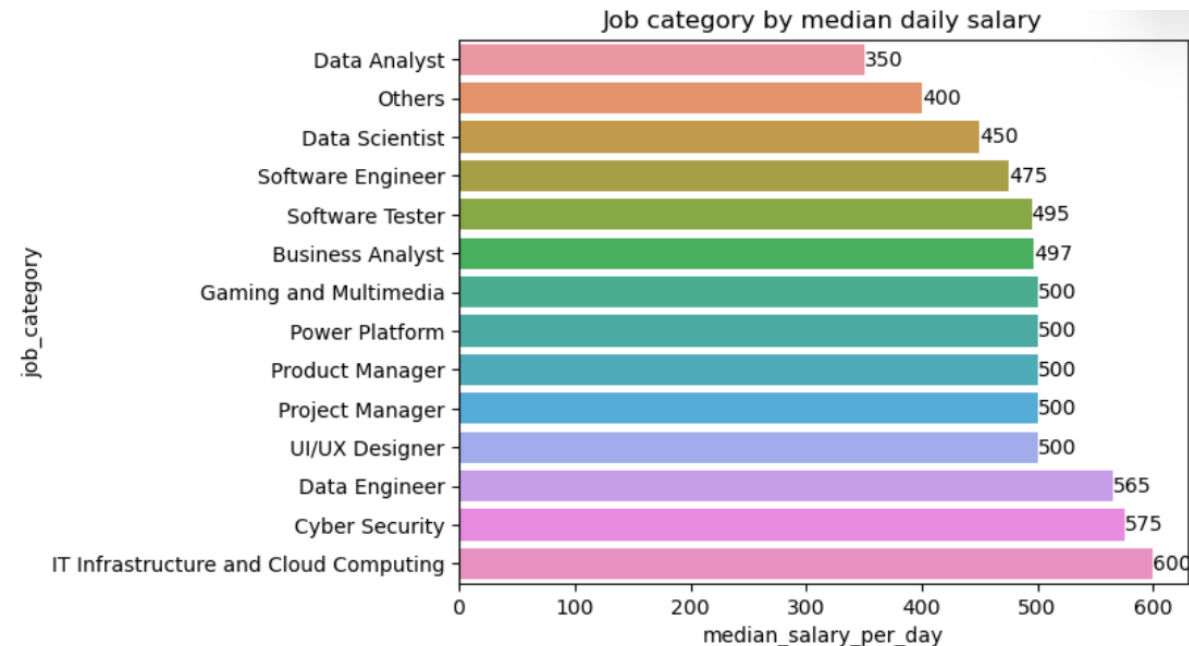


See Above

Insights: Using their Median Salary, the Data Engineer and Cloud Computing jobs advertised has the highest earning of about £60,000 each while Power Platform, Product Manager, and CyberSecurity earn about £50,000. But the least median salary is £35,000 recorded in Data Analyst career.

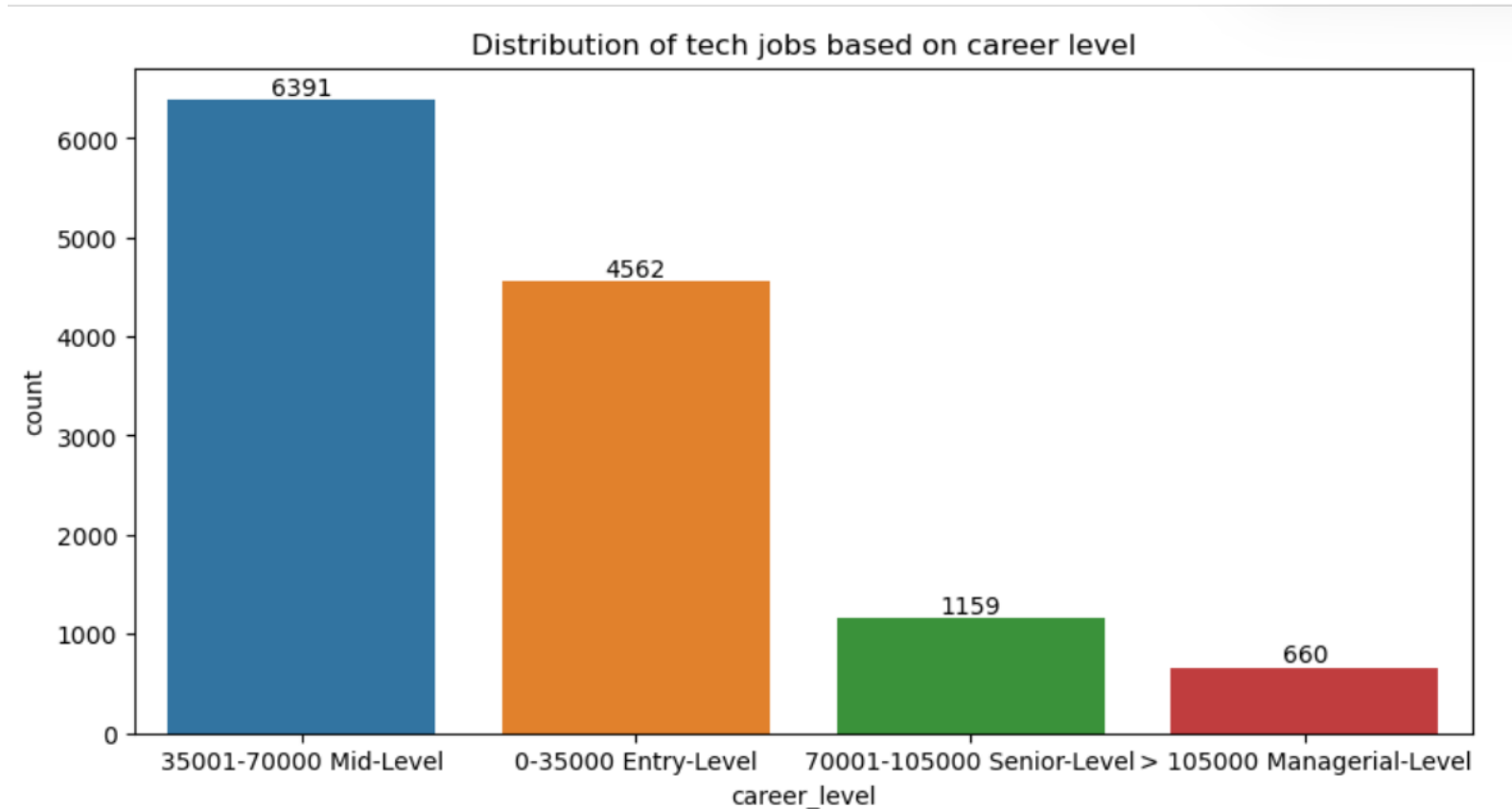
See Below

Insights; The job category with the highest (£500 to 600) median salary per day includes Cyber Security, Data Engineer, Gaming and Multimedia, Cloud computing, Power Platform, Product Manager, UI/UX Designer and Project Manager.

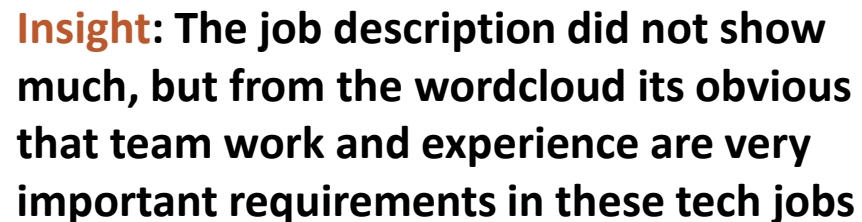




EXPLORATORY DATA ANALYSIS



Insights: About 50% of the jobs vacancies is for Mid-Level career position with salary scale within £35,001 and £70,000, while about 36% is for EntryLevel (below £35,000), and the least is 5%, which the Managerial Level career position (salary above £105,000).



Insight: The wordcloud shows that the skills in high demand are Data Analysis, Programming language and problem solving skill.





OBSERVATIONS

Based on the analysis of tech jobs from UK jobsites the following was observed:

London is the city with the highest tech job vacancy. Among the top ten companies with tech job vacancies, Oscar Associates (UK) Limited tops the chart, 2% (262) of the jobs advertised comes from them. Other companies recorded among the top ten includes Rise Technical Recruitment Limited and Noir. Software Engineer, Data Analyst, Project Manager and Business Analyst are the top 4 tech jobs being advertised in United Kingdom, while Software Tester and Power Platform are the least. About 69.5% of the jobs advertised are annual salary jobs, 5.7% are per day jobs while the least is the monthly duration job which is about 0.3%.

Half of the jobs vacancies are mid-level career position with salary scale within £35,001 and £70,000, and more than one-third are entry-level (below £35,000) position, while a minute percentage(5%), are managerial_level career position (salary above £105,000). The job description did not show much, but team work and experience are very important requirements in these tech jobs. The annual salary for some of the tech roles like Data Scientist, Product Manager, Data Engineer, Power Platform are consistent, that is they do not record any unusual values. Data Engineer and Cloud Computing jobs advertised has the highest earning of about £60,000 each while Power Platform, Product Manager, and CyberSecurity earn about £50,000. but the least median salary £35,000, is recorded in Data Analyst career. Jobs like, Data Enngineer, Software Tester, Software Engineer, Power Platform and Data Engineer are the top 5 jobs that earn between £45 to £60 on hourly basis while the least pay on hourly basis is CyberSecurity. But on a daily basis, the job category with the highest (£500 to £600) median salary pay includes Cyber Security, Data Engineer, Gaming and Multimedia, Cloud computing, Power Platform, Product Manager, UI/UX Designer and Project Manager.

The skills that are in high demand are Data Analysis, Programming languages and problem solving skill.



RECOMMENDATIONS

TO JOB SEEKERS

-For those in search of a tech job in the UK, living in the City of London or its environs, is a great advantage, as most of the vacancies are in the city of London.

-Oscar Associates(UK) is a major recruiter, therefore we advise that you submit your CV to their site and register with them if need be, for regular updates.

-Those intending to get into the tech space should ensure they are trained in institutions where post-training experience is provided as work experience is necessary in the tech space.

-If you are thinking of a career in the UK tech space, considering one of Software Engineer, Data Analyst, Project Manager or Business Analyst, is the way to go as these are the jobs in demand.

-If high annual salary is a major reason for transitioning into tech, then you may consider high paying tech jobs like Data Engineer, Cloud Computing, Power Platform, Product Manager, or CyberSecurity while the high daily pay jobs (£500 to £600) are Cyber Security, Data Engineer, Gaming and Multimedia, Cloud computing, Power Platform, Product Manager, UI/UX Designer or Project Manager.



RECOMMENDATIONS

TO EDUCATIONAL TECH PROVIDERS

- Ensure that you embed Data Analysis, Programming Languages and Problem Solving Skills into your curriculum, as they are the skills in high demand in the tech space.
- Make sure to provide some form of internship or post-training experience for the students that train in your institution, as team work and experience are key requirements for employment.
- Software Engineer, Data Analyst, Project Manager and Business Analyst are courses that are in high demand in the UK market hence should be listed as one of the taught programmes

TO EMPLOYERS

- Look into your company's salary scale along side with the median salaries in this analysis, for all level of employment and ensure your company falls within the job market trend.
- Continuous in-house training is necessary to ensure that your staff are up to date in skills like Data analysis, programming language and problem solving skills.
- From the analysis, vacancy for managerial level is few, this tells us that most companies, as a career progression path, look inwardly when filling managerial vacancies



CONCLUSION

According to Science Times, Cloud Computing (DevOps Engineer) and Product Manager were listed as one of the ten highest paying jobs in the UK, while other tech jobs listed were Data Science and Software Engineer (Science Times, 2023). Furthermore, similarities were observed when our findings were compared with a different country, in this, the US Bureau of Labour Statistics listed software developer and Data Science as one of the 20 occupations with the highest projected percentage change of employment between 2022 and 2032 (US Bureau of Labour Statistics, [September 6, 2023]).

In conclusion, our analysis aligns with the insights provided by reputable sources such as Science Times and the US Bureau of Labour Statistics. The demand for tech jobs like Cloud Computing, Product Management, Data Science, Business Analyst, Data Analyst, CyberSecurity, UI/UX Designer and Software Engineering continues to grow, making them among the highest paying and most promising tech career paths in the UK . This reaffirms the significance of pursuing a career in the tech industry, where opportunities are abundant and salaries are competitive.



CONSTRAINTS

Data quality was a major constraint in this project, there were lots of missing data and inconsistencies in the data set, limited features and, also the quantity of the data was too small. All these hindered our ability to go further in building a machine learning model.



REFERENCES

Science Times. (2023, February 23). 10 Highest Paying Tech Jobs in the UK and a Few More in High Demand. Science Times.
[<https://www.sciencetimes.com/articles/42557/20230223/10-highest-paying-tech-jobs-in-the-uk-and-a-few-more-in-high-demand.htm>]

U.S. Bureau of Labor Statistics. "Fastest Growing Occupations." Bureau of Labor Statistics, U.S. Department of Labor, modified 6 September, 2023, <https://www.bls.gov/ooh/fastest-growing.htm>



THANK YOU

FROM GROUP 7