

WEB SCRAPING OF DATA ANALYST JOBS

Presented by: **D Tadesse**

PGCE Mathematics, BSc Mathematics & Statistics: <https://github.com/datadesse>

Last updated: **18 June 2024**

London, UK

TABLE OF CONTENTS

- **Executive Summary**
- **Introduction**
- **Methodology**
- **Data Collection & Process**
- **Results**
 - **Visualization – tables, charts**
 - **Dashboard**
- **Discussion**
 - **Findings & Implications**
- **Conclusion**
- **Appendix**

EXECUTIVE SUMMARY

In this analysis I will explain why:

- Tuesday is the best day of the week to use the tool to scrape the website
- £39,132 is the average annual salary of a Data Analyst in England
- Business Intelligence (BI) Data Analyst is the top paid role
- ITonlinelearning Recruitment is the most popular employer
- Competitive and Fixed Salaries are more appealing to Data Analysts than other salaries
- The majority of the jobs offer Permanent full-time working hours

INTRODUCTION

Objectives of this analysis

The recruitment agency where I work as a data analyst has asked me to create a web scraping tool that can automatically extract details of Data Analyst jobs in England from a job posting site & make recommendations from my findings.

PROBLEMS TO SOLVE

- Which employer advertised the most jobs ?
- When is the best time to use the scrapping tool ?
- What are the highest and least paid roles ?
- How much is the average salary ?
- Is there any other findings useful for the agency to consider ?

METHODOLOGY

Data source:

1. I browsed through job websites for Data Analyst role in England and picked www.reed.co.uk
2. **BeautifulSoup, Requests and Pandas** are used to scrape details of Data Analyst jobs advertising- titles, dates, salaries, locations, employers and working hours to answer the problems
3. The title column helps me to filter out jobs only containing the Data Analyst roles
4. The salary column allows me to calculate average, median, min and max of figured salaries and investigate other salaries
5. The employer column is crucial for the agency to target partnership work, to stay commutative, increase the efficiency & quality of job vacancy sourcing.

DATA COLLECTION & PROCESS

- Experimented with my codes on the first page of the website then created a single function that automatically extracts all pages to download 523 rows as a CSV file format
- From 523 rows duplicates of 229 rows removed and only 294 rows used for my analysis
- After outliers removed the min_salary & max_salary columns contain only 160 & 163 rows respectively.

TOOLS USED:

- Python's libraries- **BeautifulSoup, Requests & Pandas** used to scrape the data , filter out titles, calculate the median and average of salaries, and identify five top employers, highest and least paid roles.
- **Matplotlib, Seaborn and Tableau Public** used to create tables, charts and dashboard to visualise the results.

Fig 1

Top Five Data Analyst Employers in England

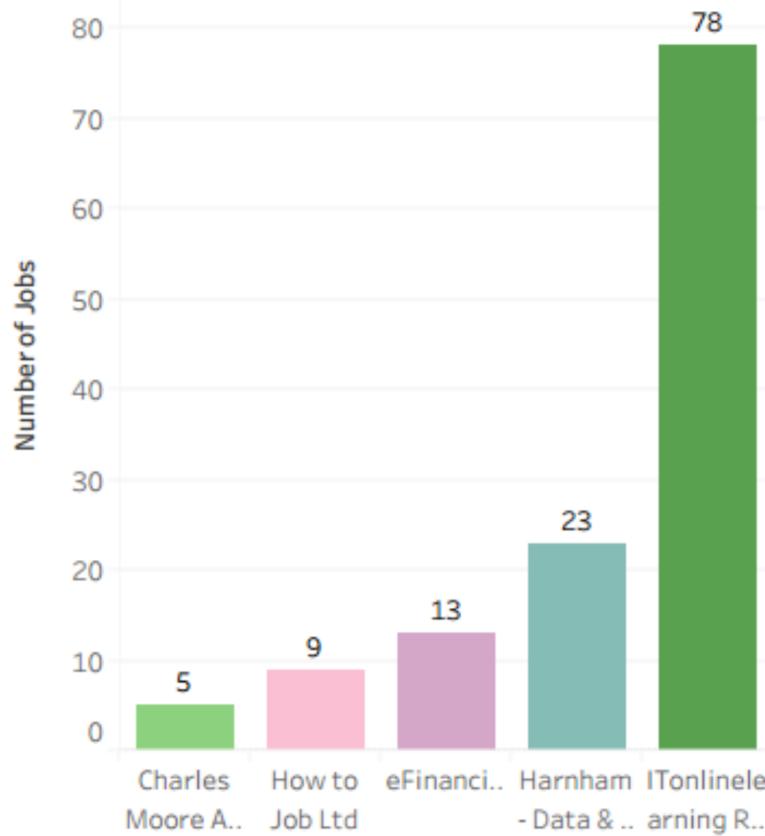
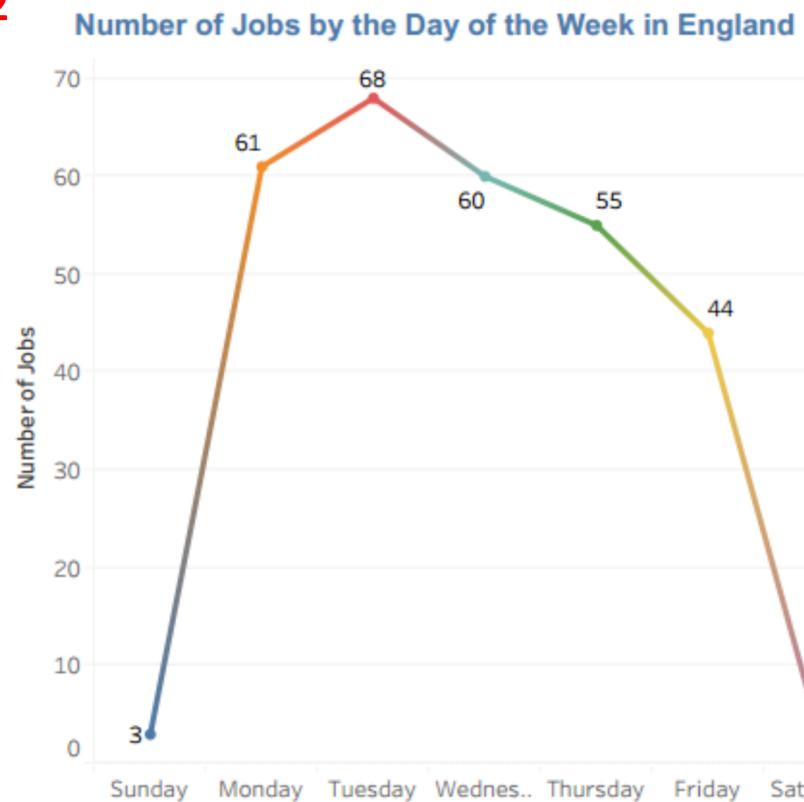


Fig1 reveals that **ITonlinelearning Recruitment** has a significant **(26.5%)** number of jobs advertised suggesting that it is the most popular employer than others and indicates that the agency would benefit from partnership work with **ITonlinelearning Recruitment**.

Fig 2



As we can see in **Fig 2** the advertising rapidly increases from Monday reaching the maximum on **Tuesday**, decreasing gradually from Wednesday and dropping sharply on the weekend. This result indicates that Tuesday has the highest number of jobs advertised suggesting that the agency should consider applying the scraping tool using all its resources on **Tuesdays** to get the relevant openings to their clients.

Fig 3

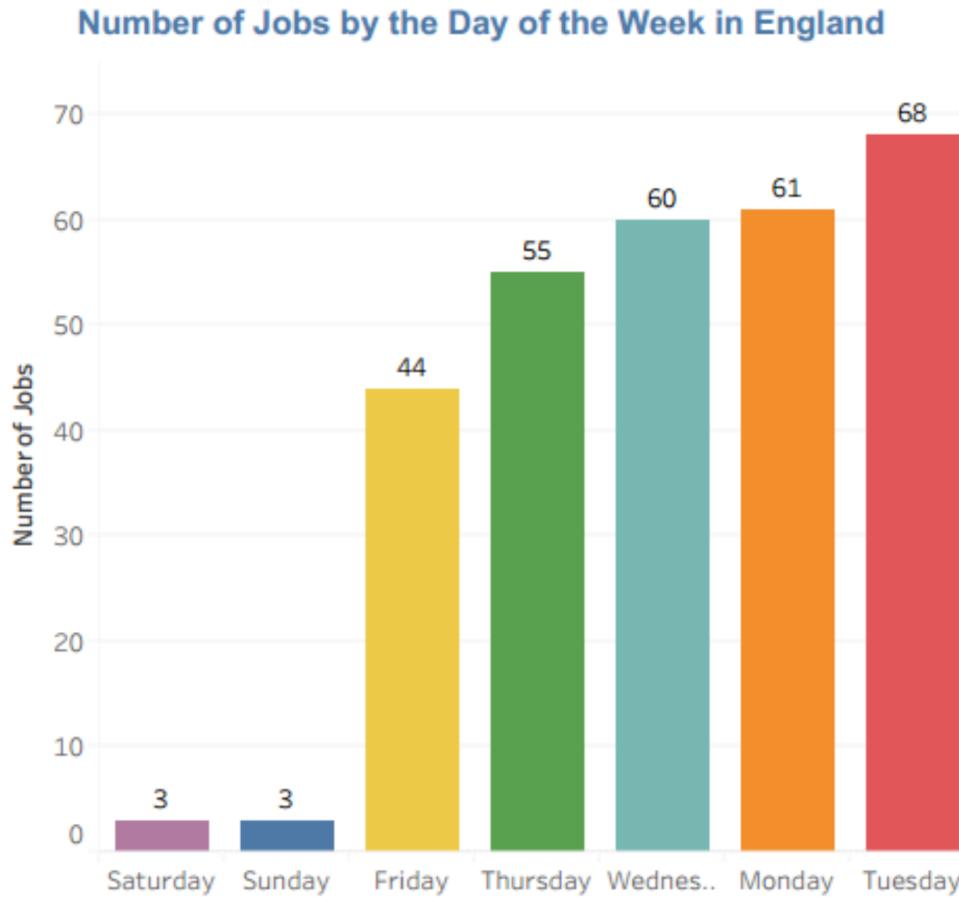


Fig 3 shows that **Tuesday** is the **highest** implying that **23%** of the jobs advertising are on **Tuesdays**. This finding indicate that the agency would benefit the most if they use full resources of the scraping tool on **Tuesdays** than other days

Fig 4

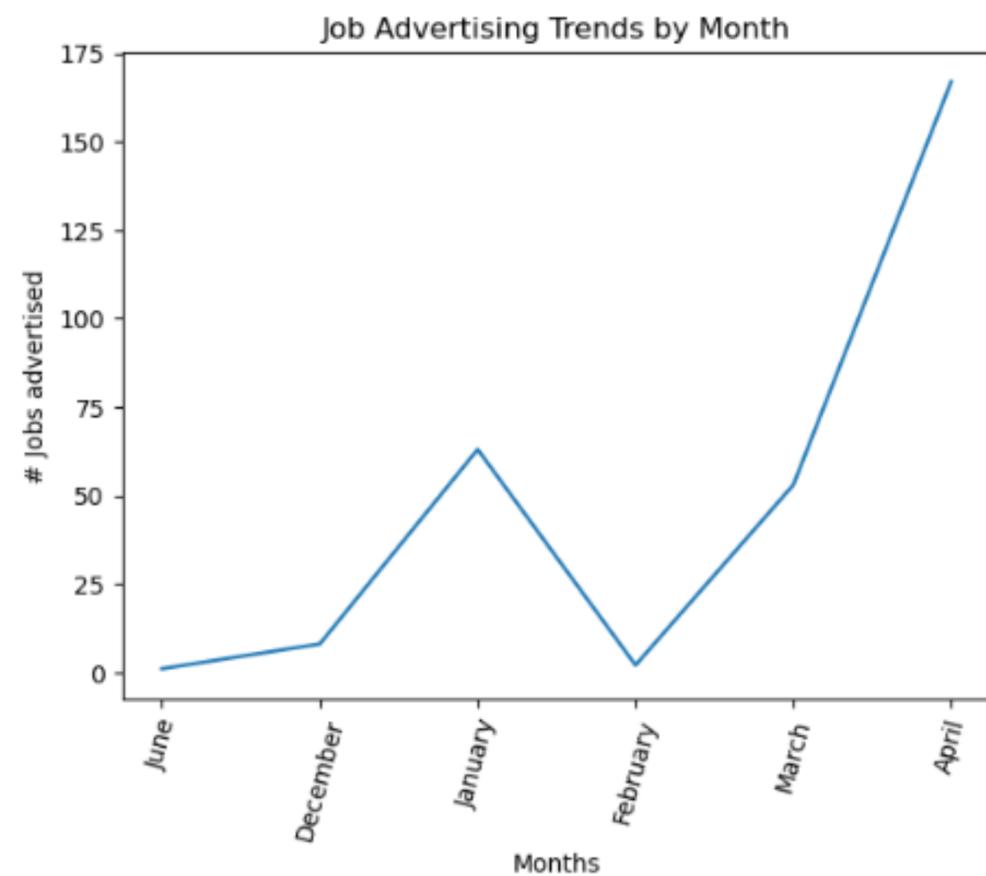


Fig 4 shows that the advertising started increasing from December 2023 to January 2024, dropped sharply in February then increased from March reaching the highest in **April** (167). Data is missing between June & December 2023 so the month is not an appropriate measure hence omitted from this analysis

Top 5 Highly Paid Data Analyst Jobs in England

Fig 5

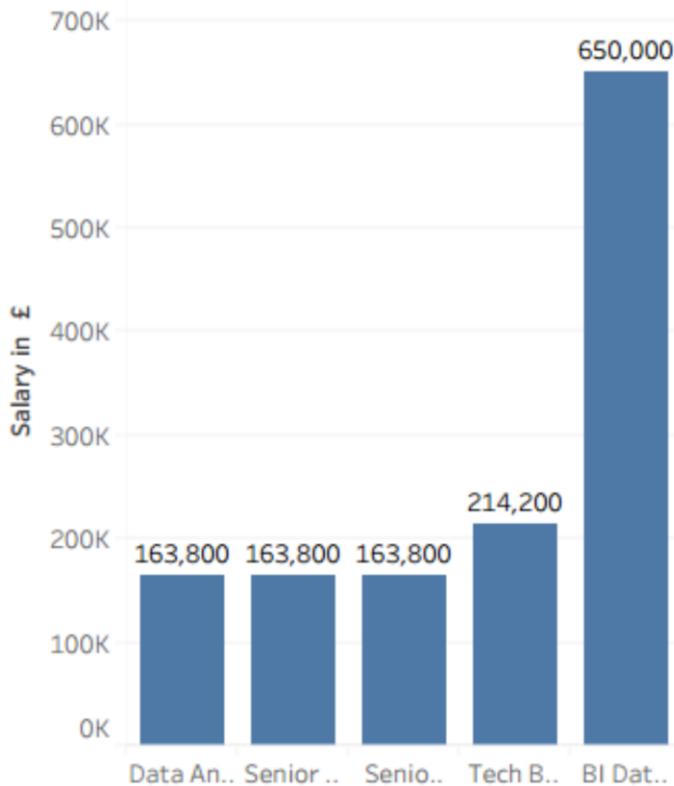


Fig 5 is evidence that the **Business Intelligence (BI) Data Analyst** role earns **£650,000 per year**. This finding implies that this is a significant amount of money hence it is the top paid role. The agency should consider reaching out to **BI Data Analysts** to stand out from its competitors

FIVE TOP PAID DATA ANALYST ROLES IN ENGLAND

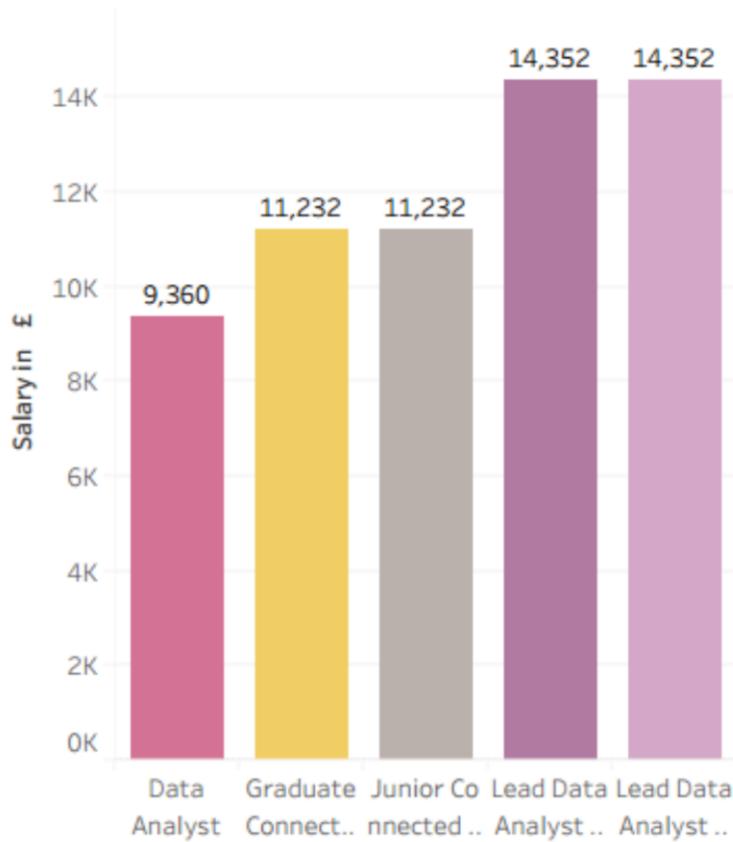
Fig 6

	title	salary	max_salary
191	BI Data Analyst	£55,000 - £650,000 per annum	650000
272	Tech BA/Data Analyst - Booking Model Controls/...	£700 - £850 per day	214200
273	Senior Business/Data Analyst - Sql, Excel, Dat...	£550 - £650 per day	163800
173	Senior Data Analyst	£550 - £650 per day	163800
216	Data Analyst - Markets Transformation	£600 - £650 per day	163800

Fig 6 is evidence that BI Data Analyst, Tech BA/Data Analyst Booking Model Controls/Tableau, Senior Business/Data Analyst, Senior Data Analyst & Data Analyst Markets Transformation are the top five paid roles in their increasing order suggesting that the agency would benefit and stands out from reaching out to these professionals.

Five Least Paid Data Analysts Jobs in England

Fig 7



As we can see in Fig 7, the least paid Data Analyst earns £9360 per year suggesting that this is the smallest amount paid. The agency will stand out from its competitors by setting the **minimum annual salary to more than £9360**.

FIVE LEAST PAID DATA ANALYST ROLES IN ENGLAND

Fig 8

	title	salary	min_salary
128	Data Analyst	£15 - £17 per hour	9360
95	Data Analyst	£18.71 per hour	11232
237	Graduate Connected Car Data Analyst	£18.01 per hour	11232
241	Junior Connected Car Data Analyst	£18.01 - £22.67 per hour	11232
270	Lead Data Analyst Workforce - NHS (b7)	£23 - £26 per hour	14352

Based on **Fig 8** a Data Analyst can earn a starting salary of as little as £15 per hour implying that this is the least paid the agency should consider setting the floor salary to more than £15 per hour to stand out from its competitors.

AVERAGE ANNUAL SALARY

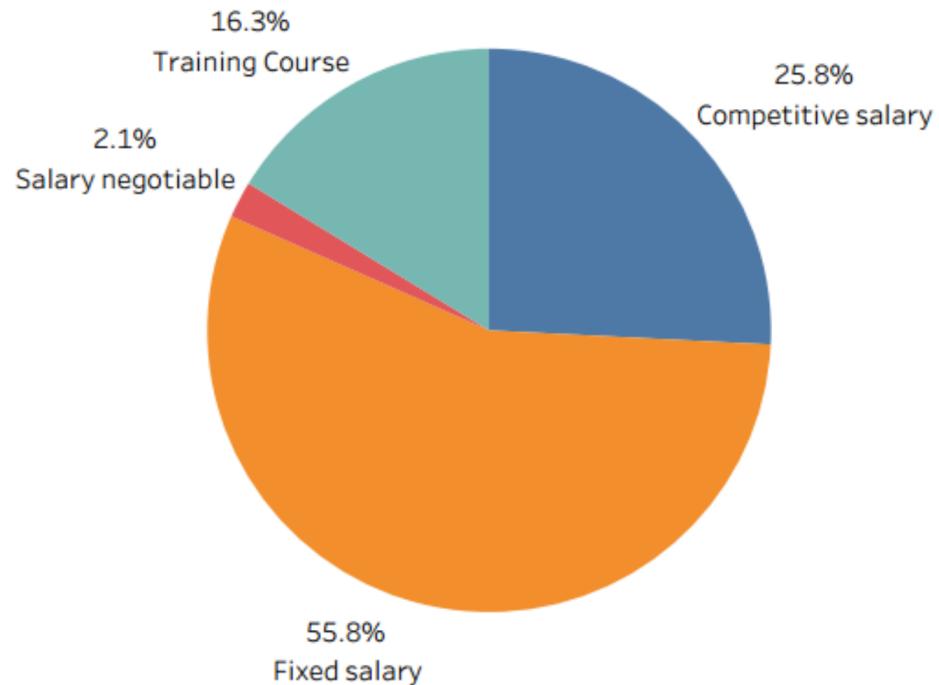
Fig 9

	min_salary	max_salary	average salary
Median	£35,000	£40,000	£37,500
Mean	£37,946	£40,318	£39,132

Fig 9 shows that £39,132 is the average annual salary of a Data Analyst using the mean as the best measure indicating that my finding lies between uk.talent.com £35000 & close to the low range of www.reed.co.uk £40,150. The agency should consider using £39,132 as a guide for the annual salary

TYPES OF SALARIES

Fig 10



From **Fig 10** we learn that **Fixed Salary 55.8% & Competitive 25.8%** are the top two salaries indicating that they take **81.6%** which is a significantly higher proportion of the salaries and they are more popular than others. The agency should consider offering **Competitive & Fixed salaries** to appeal to more data analysts.

WORKING HOURS

Fig 11

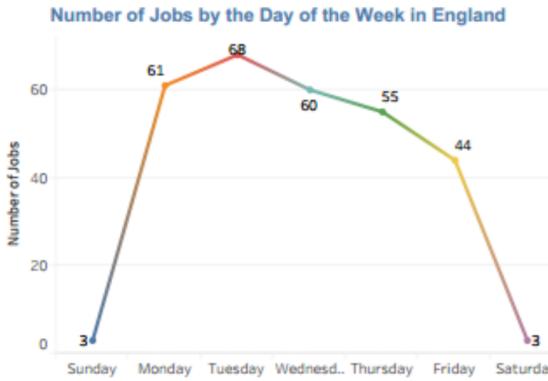
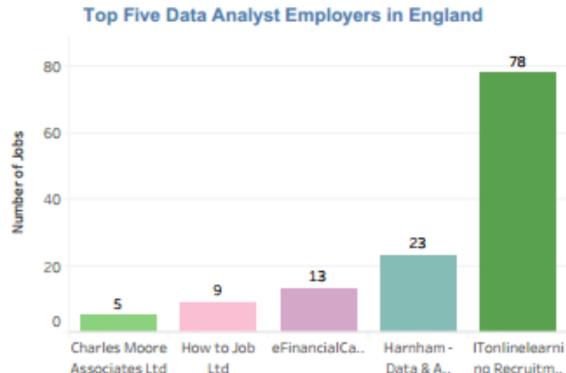
	working_hours	# of jobs	%
0	Permanent, full-time	255	86.7%
1	Contract, full-time	29	9.9%
2	Temporary, full-time	6	2.0%
3	Permanent, full-time or part-time	4	1.4%

Fig 11 shows that a large number of jobs **86.7%** offer **Permanent full-time** suggesting that it is the biggest proportion and most popular working hours than others. The agency should consider offering **Permanent full-time** roles to stay competitive

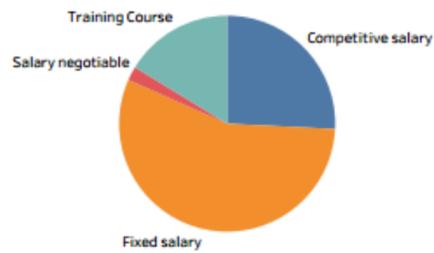
TABLEAU DASHBOARD

Fig 12

Analysis of Data Analyst Job Advertising in England



Jobs with Fixed and other Salaries in England



Top 5 Highly Paid Data Analyst Jobs in England

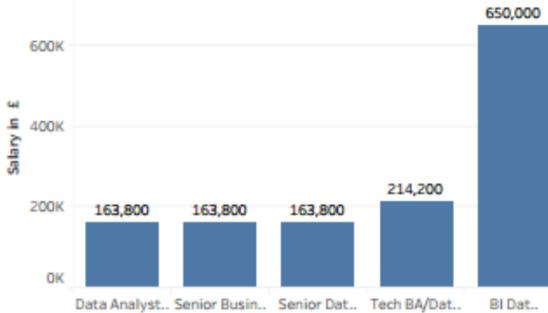


Fig 12 shows the number of jobs by the day of the week, salary types, the top five employers and highly paid Data Analyst roles using Tableau Public

OVERALL FINDINGS & IMPLICATIONS

Findings	Implications
Tuesday has the highest number of jobs advertised (Fig 2 & Fig 3)	The agency will get the relevant openings to their clients if it uses the scraping tool on Tuesdays
81.6% the jobs are Fixed & Competitive salaries and are very popular (Fig 10)	The agency should consider offering Competitive & Fixed salaries to appeal to more data analyst
Business Intelligence(BI) Data Analyst role earns significant amount of money £650K (Fig 5 & Fig 6)	The agency should consider reaching out to BI Data Analysts
The smallest salary is £15 per hour or £9360 per year (Fig 7 & Fig 8)	The agency would be competitive from setting the floor salary to more than £15 per hour or £9360 per year

Continuous on the next page ...

OVERALL FINDINGS & IMPLICATIONS continued from the previous page

Findings	Implications
<p>£39,132 is the annual average salary (Fig 9)</p>	<p>The agency should consider to use £39,132 as a guide for annual salary</p>
<p>ITonlinelearning Recruitment offers the highest number of jobs advertised and it is the most popular employer (Fig 1)</p>	<p>The agency should consider partnership work with ITonlinelearning Recruitment</p>
<p>Permanent full-time working hour takes the largest proportion 86.7% from all jobs advertised (Fig 11)</p>	<p>The agency will stand out from its competitors if it considers offering more Permanent full-time roles</p>

SUMMARY

- I used the first page of the website to extract titles, salaries, locations, links, and descriptions of each job then experimented with the code before applying it to all pages
- Created a single function that can automatically extract the required information from all pages and save the data as a CSV file
- Processed the data then analysed using Python's data visualization libraries- Matplotlib and Seaborn to make comparisons between employers, salaries, working hours and date – to see the change over time and when to use the scrapping tool
- From 523 rows of the scrapped data, 229 (44%) duplicates were removed and only 294 rows (56%) were used for my analysis
- Outliers in the min_salary and max_salary columns were significant and removed before the analysis
- Max_salary column after cleaned from outliers & contains 55 % (163) of 294 rows, 31% from the original 523 rows
- Min_salary column after cleaned from outliers & contains 54 % (160) of 294 rows, 30.5% from the original 523 rows
- The dataset I used for min and max salary analysis is just over half of 294 rows

CONCLUSION

As I motioned in the EXECUTIVE SUMMARY section my analysis findings confirm that:

- The average annual salary of a **Data Analyst in England** is £39,132 using the mean of min and max columns as a measure
My finding falls between uk.talent.com (£35000) and close to the low range of www.reed.co.uk (£40,150)
- Tuesday has the highest number of jobs advertised and it is the best day of the week for the agency to use the scrapping tool
- A Business Intelligence (BI) Data Analyst earns a significant amount of money **£650,000 per year**
- ITonlinelearning Recruitment is the most popular employer
- Competitive and Fixed Salaries are more appealing to Data Analysts than other salaries
- The majority of the jobs offer **Permanent full-time working hours**
- The **smallest salary is £15 per hour or £9360 per year**

continues on the next page...

CONCLUSION continued from the previous page

Limitations

- The data is very small and doesn't represent population data
- The jobs appear on the website for a limited period of time until someone is hired so this may have impacted the 2023 data as well as my analysis so my conclusions drawn here may not reflect the real world scenario.

Suggestions

- Further investigation is required if these jobs advertised are seasonal or companies demand is a one off or continues.
- Some jobs are advertised by recruitment agencies on behalf of employers so there may be more jobs on the employers' website
- On the day of scrapping (28/04/2024) the website had 2872 Data Analyst Jobs in England but my program extracted 2874 which is an extra of 2 jobs. A further investigation is required as to why this happened.

Appendix

Data source

- The data sample size 294 rows is a very small and contains a few months from 2023
- This analysis is just to demonstrate that I can apply similar methods to large scale of data sets

My Tableau Public charts & Dashboard link:

<https://public.tableau.com/app/profile/d.tadesse/vizzes>

References:

- Getting Started with Data Science by Murtaza Haider (book)
- [ChatGPT](#) & [Stack Overflow](#)

Acknowledgements to:

- [Google](#) & [Coursera](#) for providing the [Google Data Analyst Professional Certificate](#) course.

Thank You