Indeed.com Job Listings Analysis

Anouk Heemskerk, Georgiana Huţanu, Renée Nieuwkoop and Alan Rijnders (2021)

1. Introduction

The job market is becoming ever more competitive, as a student nowadays it is harder and harder to land the job of your dreams. COVID-19 has even made it more challenging for recent graduates to find a fitting role after graduation. A lot of students complain about the gap between skills learned during their studies and the actual skills required in the workfield by companies. This motivates our investigation into Indeed job vacancy postings to find the skills that are actually required for different job types.

As Marketing Analytics students, we have both the marketing and analytical knowledge to succeed in a range of different job types in the field of marketing. We are interested in what the actual required skills are for jobs in fields related to our studies. Becoming a data scientist, data analist or maybe marketeer after our studies will require different skills and competencies and we would like to know to which degree we possess these skills and to which extent our study program effectively prepares us for the job market.

We aspire to find insights to help ourselves as well as our fellow students in making a choice relating to which skills to improve and perhaps which skills to forego in order to effectively land their first job. We aim to conduct our investigation in such a way that the methodology can be used by anyone for any location in the world and for any job type on Indeed.com. First of all, for our fellow students, the keyword and location analysis provide easily interpretable tables to see which skills are in demand for each roles, together with the best locations for jobs in the Netherlands. Our entire project is accessible and usable. The entire workflow can be tweaked to your needs by simply changing the search term in the scraper. This makes our project valuable for all job seekers in the world, as they can reproduce our data scraping and analysis for their specific job wishes.

For this project we decided to narrow down our investigation to the Netherlands. Three of our four members originate from the Netherlands and since we and our classmates are all studying at Tilburg University in the Netherlands right now, job options in the Netherlands are most relevant to us. Furthermore, we decided to investigate the 4 jobs most closely related to our master in Marketing Analytics program, being data scientist, marketeer, data analist and marketing analist. By narrowing down the project by not including too many locations and job searches, the project workflow will be much smoother and easier to reproduce for anyone interested in doing so.

2. Methodology

First, we built a web scraper that scrapes the vital information of each job posting associated with a specific job search. We used the BeautifulSoup package to collect job ids, job titles, salary, company name, dates, job summary and salaries if available. Afterwards, we scraped the job descriptions of the same search results to obtain the job descriptions of each job posting in a separate dataset by using a chrome webdriver with Python Selenium. Due to restrictions with captcha's, we collected multiple small batches of datasets per search term. In R we merged the data into one big file per search term by joining the datasets on the unique job id that serves as an identifier for each separate job advert on Indeed. After merging the files into one dataset per search term, we cleaned the data by removing duplicate entries and by cleaning up messy string location names. Location names such as Amsterdam-Zuid or Velsen-Noord were unwanted because they would appear as two distinct location in our analysis. Therefore, we wrote a function that removed all such extra unnecessary information in the location strings. We added a function that replaced certain unspecified

locations such as Nederland or Randstad into Unknown, signifying that a job did not specify a specific location.

The final step in the cleaning process was cleaning the salary data. The salary data was quite sparse and also very messy. Roughly three-quarters of the data was missing, and for the data available, different measures were given. Some jobs gave hourly salary rates, whereas other jobs utilized monthly or yearly income as a salary measure. We made a function that removes unnecessary character strings in the salary data and converted all different salary measures into yearly income as a standard. We calculated yearly income for hourly rates based on 40 hour work week. We multiplied the hourly rate * 40 * 4 * 12 to get a yearly income rate based on the hourly wage. For the monthly income, we simply multiplied with 12 to arrive at a yearly income. Quite a few jobs gave a salary range instead of a fixed number. For those jobs, we decided to take the middle of the range. We kept the observations with missing values for salary in the dataset for the first part of the analysis because it would significantly reduce the number of observations otherwise.

3. Overview of the data

After downloading and merging data we have four datasets in total. One dataset per search term for the following search terms: Data Scientist, Data Analist, Marketing Analist and Marketeer. The following section will provide a short overview of how the datasets look like.

3.1 This is what the data scientist dataset looks like

We can see that there are 706 job listings. It contains 12 columns.

```
##  [1] "id"                 "title"                "company"
##  [4] "location"           "postingdate"          "today"
##  [7] "summary"            "salary"               "url"
## [10] "description"        "scrapetimesdescription" "salary_good"
```

Here, a summary of the data scientist dataset be observed.

```
##       id               title              company             location
##  Length:706         Length:706         Length:706         Length:706
##  Class :character   Class :character   Class :character   Class :character
##  Mode  :character   Mode  :character   Mode  :character   Mode  :character
##
##
##
##
##  postingdate           today              summary             salary
##  Length:706         Length:706         Length:706         Length:706
##  Class :character   Class :character   Class :character   Class :character
##  Mode  :character   Mode  :character   Mode  :character   Mode  :character
##
##
##
##
##      url             description        scrapetimesdescription  salary_good
##  Length:706         Length:706         Length:706              Min.   :   4800
##  Class :character   Class :character   Class :character        1st Qu.:  36600
##  Mode  :character   Mode  :character   Mode  :character        Median :  49908
##                                                                Mean   :  78627
##                                                                3rd Qu.:  64800
##                                                                Max.   :2098560
##                                                                NA's   :557
```

3.2 This is what the data analist dataset looks like:

We can see that there are 825 job listings. It contains 12 columns.

```
##  [1] "id"                 "title"                "company"
##  [4] "location"           "postingdate"          "today"
##  [7] "summary"            "salary"               "url"
## [10] "description"        "scrapetimesdescription" "salary_good"
```

Here, a summary of the data analist dataset can be observed.

```
##       id              title            company            location
##  Length:825        Length:825        Length:825        Length:825
##  Class :character  Class :character  Class :character  Class :character
##  Mode  :character  Mode  :character  Mode  :character  Mode  :character
##
##
##
##
##   postingdate         today            summary            salary
##  Length:825        Length:825        Length:825        Length:825
##  Class :character  Class :character  Class :character  Class :character
##  Mode  :character  Mode  :character  Mode  :character  Mode  :character
##
##
##
##
##      url            description      scrapetimesdescription  salary_good
##  Length:825        Length:825        Length:825        Min.   :   6000
##  Class :character  Class :character  Class :character  1st Qu.:  34506
##  Mode  :character  Mode  :character  Mode  :character  Median :  44412
##                                                        Mean   :  62025
##                                                        3rd Qu.:  54054
##                                                        Max.   :3432960
##                                                        NA's   :598
```

3.3 This is what the marketing analist dataset looks like:

We can see that there are 409 job listings. It contains 12 columns.

```
##  [1] "id"                 "title"                "company"
##  [4] "location"           "postingdate"          "today"
##  [7] "summary"            "salary"               "url"
## [10] "description"        "scrapetimesdescription" "salary_good"
```

Here, a summary of the data analist dataset can be observed.

```
##       id              title            company            location
##  Length:409        Length:409        Length:409        Length:409
##  Class :character  Class :character  Class :character  Class :character
##  Mode  :character  Mode  :character  Mode  :character  Mode  :character
##
##
##
##
##   postingdate         today            summary            salary
##  Length:409        Length:409        Length:409        Length:409
##  Class :character  Class :character  Class :character  Class :character
```

```
##   Mode  :character   Mode  :character   Mode  :character   Mode  :character
##
##
##
##
##      url             description       scrapetimesdescription  salary_good
##   Length:409          Length:409          Length:409          Min.   : 4200
##   Class :character    Class :character    Class :character    1st Qu.:36525
##   Mode  :character    Mode  :character    Mode  :character    Median :47100
##                                                               Mean   :46376
##                                                               3rd Qu.:55392
##                                                               Max.   :96000
##                                                               NA's   :345
```

3.4 This is what the marketeer data set looks like:

We can see that there are 872 job listings. It contains 12 columns.

```
##  [1] "id"                 "title"                  "company"
##  [4] "location"           "postingdate"            "today"
##  [7] "summary"            "salary"                 "url"
## [10] "description"        "scrapetimesdescription" "salary_good"
```

Here, a summary of the marketeer dataset can be observed.

```
##      id                title              company             location
##   Length:872          Length:872          Length:872          Length:872
##   Class :character    Class :character    Class :character    Class :character
##   Mode  :character    Mode  :character    Mode  :character    Mode  :character
##
##
##
##
##   postingdate           today              summary              salary
##   Length:872          Length:872          Length:872          Length:872
##   Class :character    Class :character    Class :character    Class :character
##   Mode  :character    Mode  :character    Mode  :character    Mode  :character
##
##
##
##
##      url             description       scrapetimesdescription  salary_good
##   Length:872          Length:872          Length:872          Min.   :   8400
##   Class :character    Class :character    Class :character    1st Qu.:  30600
##   Mode  :character    Mode  :character    Mode  :character    Median :  36000
##                                                               Mean   :  58519
##                                                               3rd Qu.:  42000
##                                                               Max.   :2400000
##                                                               NA's   :653
```

4. Keyword Analysis

In the following section we provide the analysis for our research. We performed three types of analyses.

1. Skill keyword analysis: we analyze the frequency of occurrence for certain technical skills related to marketing analytic.

2. Location analysis, after knowing the right skills to pursue for the next job we provide a location analysis in which we investigate what the top locations for our four jobs.

3. Salary analysis: we finish with a short analysis of average salaries for each job and the best locations salary wise for each job.

4.1 Getting the skill counts for the most important skills in data science jobs

Counts per skill for Data Science job listings
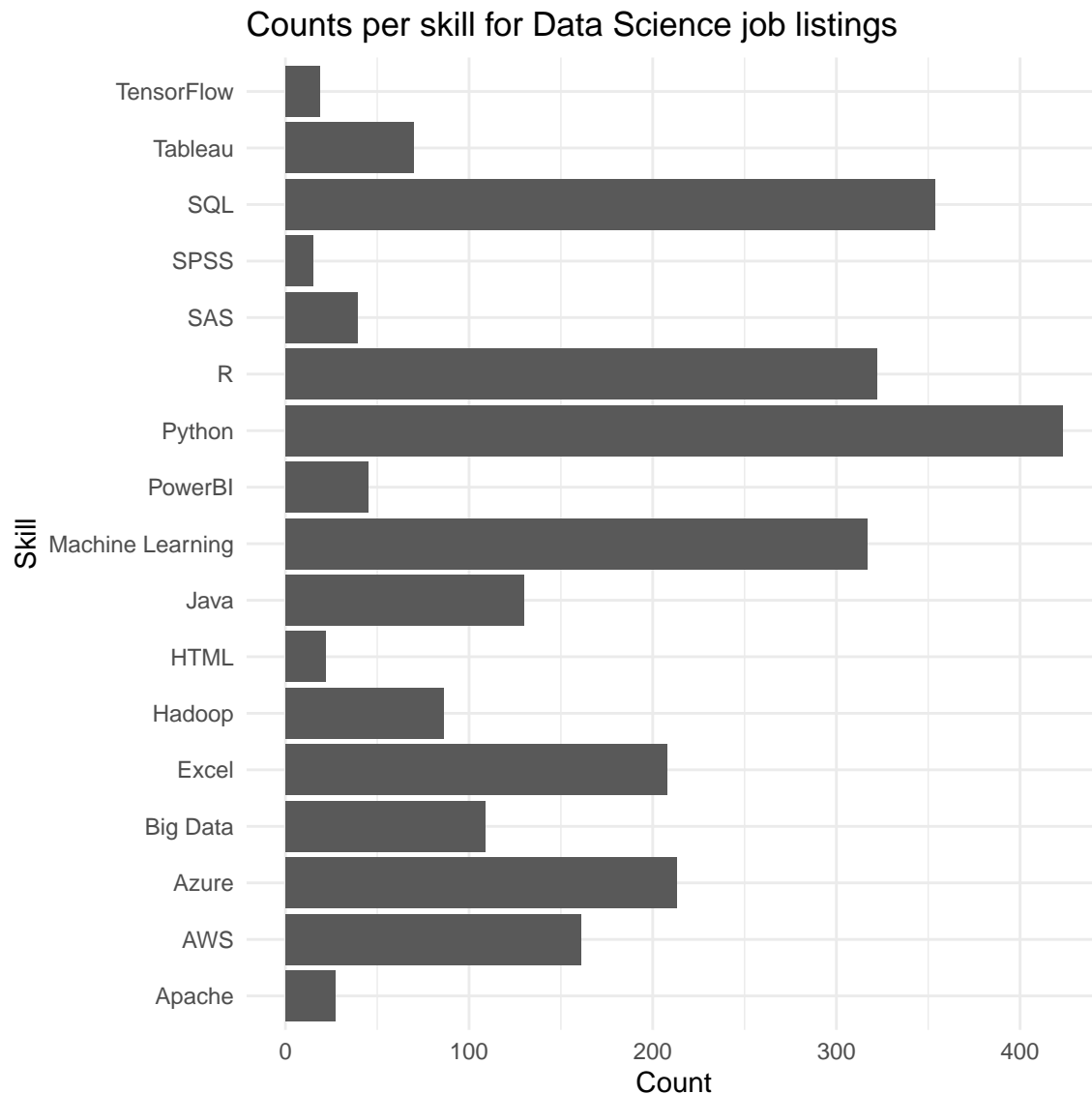
This is a plot of the core technical skills and their counts in the data scientist job listings. As we can see, there are four clear top skills to learn to become a data scientist. These include Python, SQL, R and machine learning, with each having counts over 300. Other important skills seem to be Java, Excel, Azure and AWS (Amazon Web Services), with each having counts over 100.

However, counts do not paint the entire picture. Some job listings have double or even more counts of the same skill within the same description. To control for this, we also calculate the percentage of job listings wherein each skill is featured in the description. For example, if the skill R is mentioned in half of the job listings, it will have a value of 0.5.

Percentage per skill for Data Science job listings
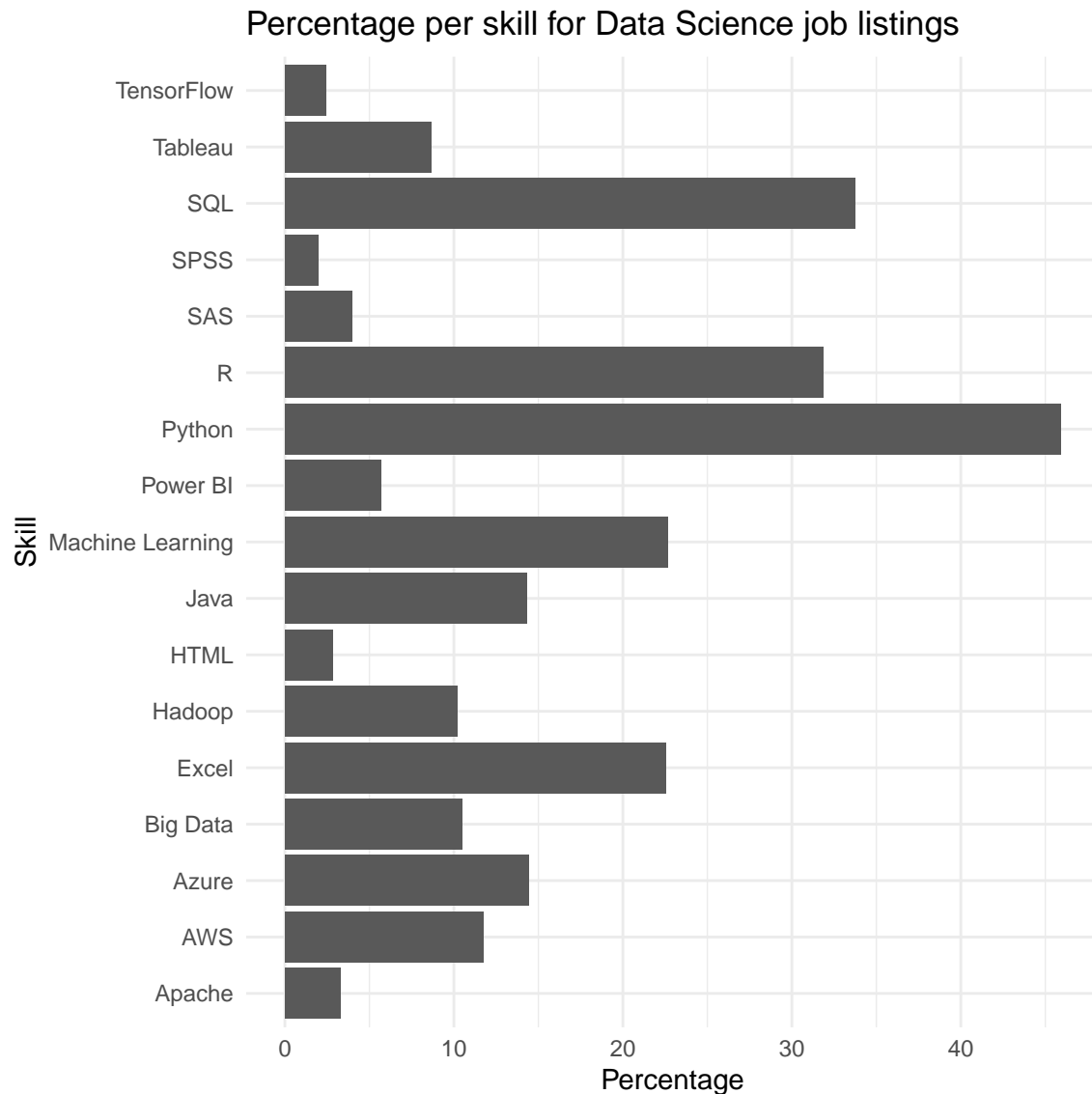
The percentage plot shows us that the percentage analysis bears quite similar results to the count analysis. Again, Python, R, machine learning and SQL are the most important skills. However, the difference between Python and the other skills became bigger in terms of percentage occurrence compared to count. Also in percentage terms, Excel is becoming more important, catching up with the top four skills and being almost equally important as machine learning.

4.2 Getting the skill counts for the most important skills in data analist jobs

Counts per skill for Data Analist job listings
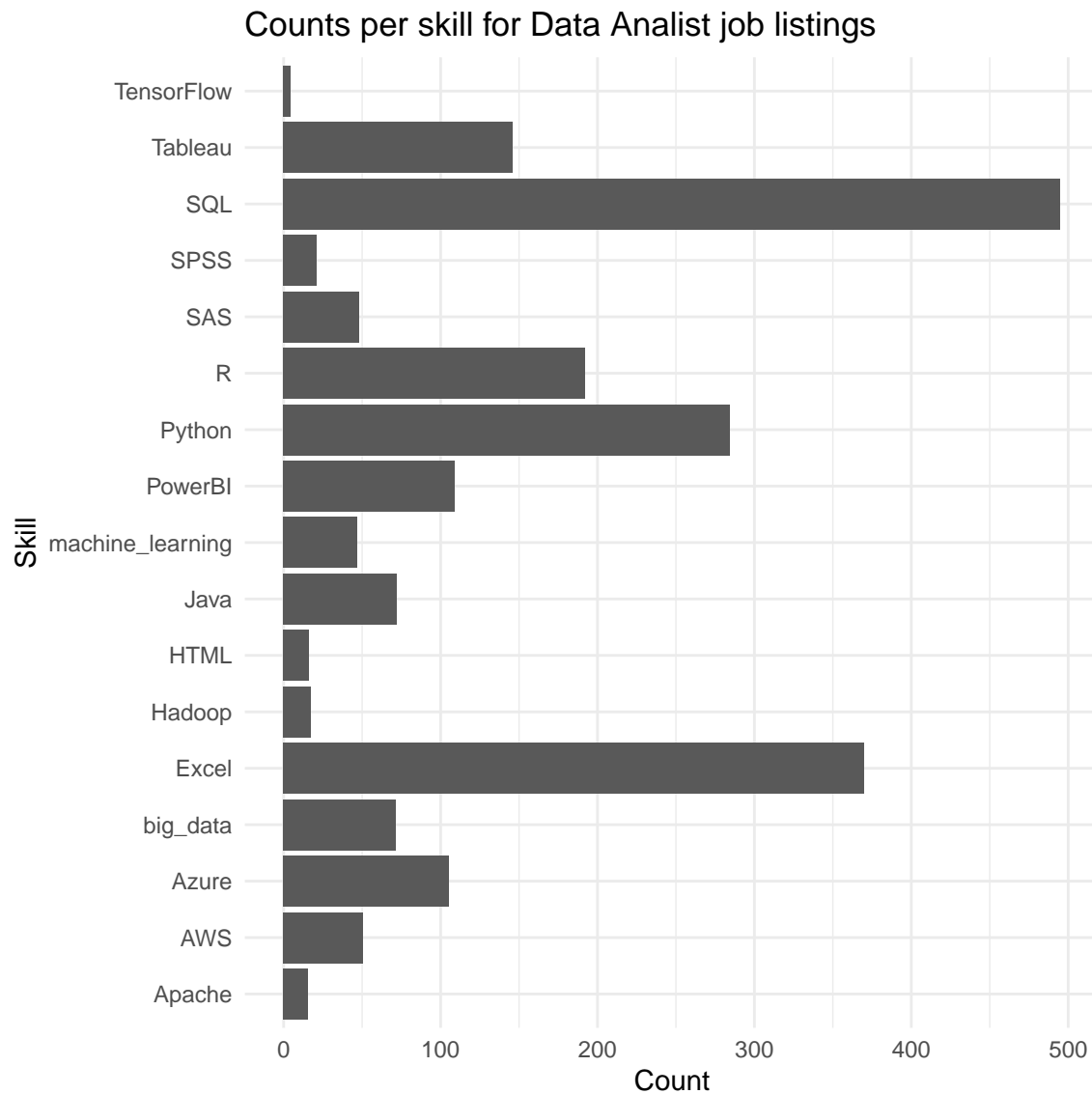
As we can see, quite a different picture for data analist jobs compared to the data science jobs. The clear winner is SQL, which approaches a count of 500. After SQL, Excel is the second most sought after skill with over 350 counts. In third and fourth place come the two programming languages Python and R, with Python slightly edging R in terms of importance. Other important skills include Tableau, PowerBI and Azure.

Percentage per skill for Data Analist job listings
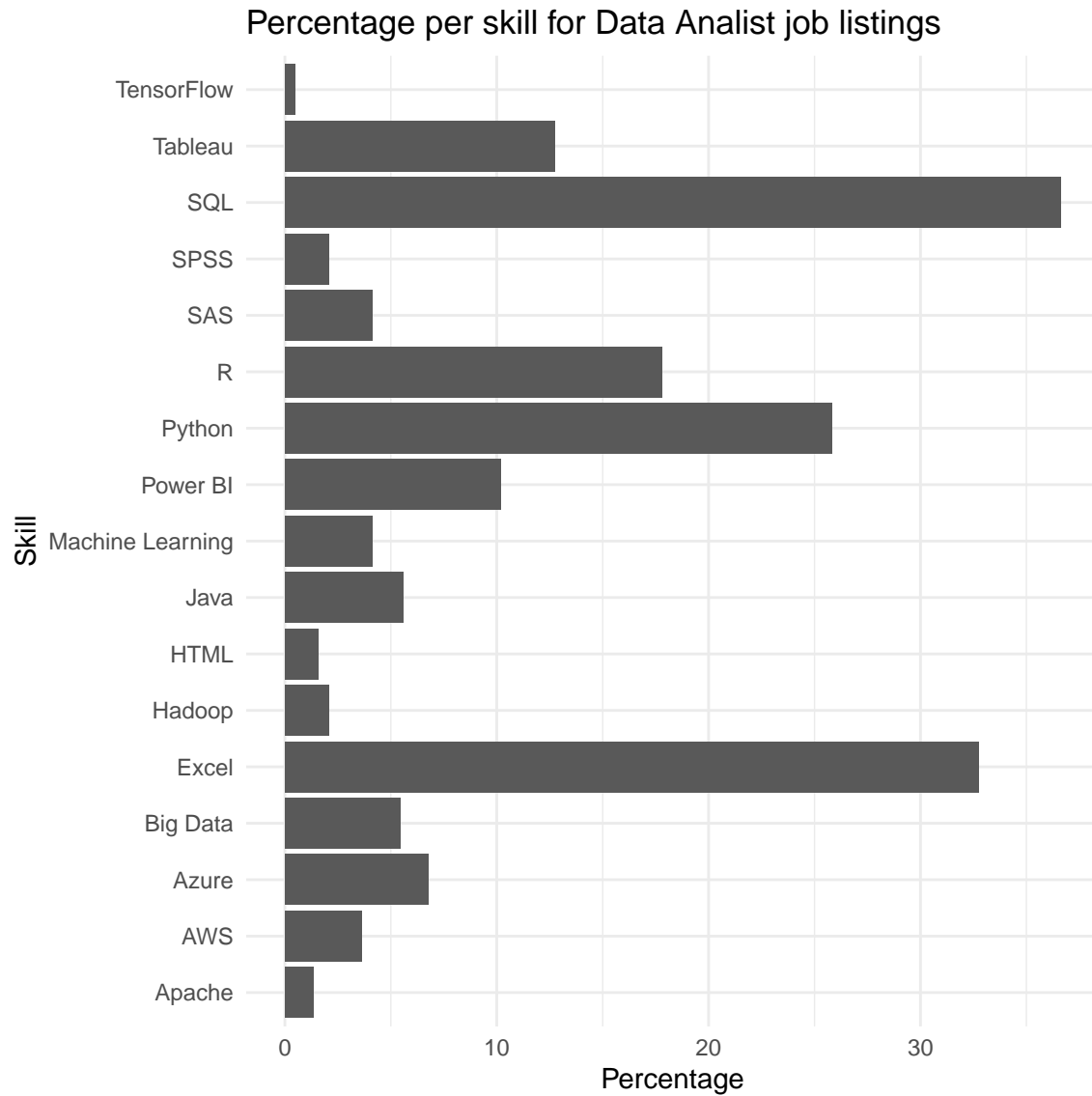
Again similar results for the percentage analysis. However, the difference between SQL and Excel is declining, with both being almost equally important. Also the difference with number three, Python, seems slightly smaller than in the count analysis. Again, other important skills are R, PowerBI and Tableau.

4.3 Getting the skill counts for the most important skills in marketing analist jobs

Counts per skill for Marketing Analist job listings
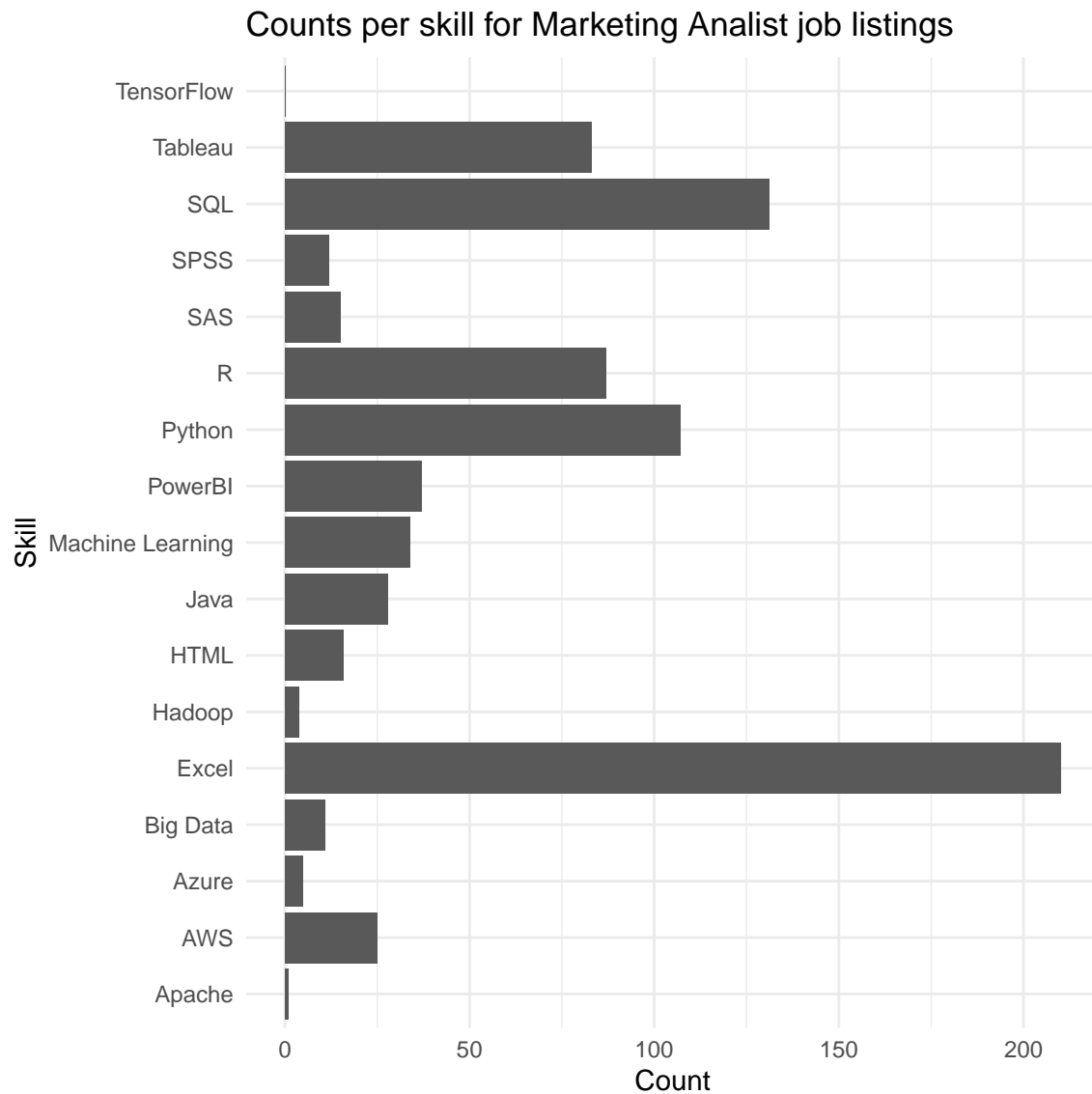
The clear winner for skills to learn for a marketing analist is Excel, being the only skill reaching a count over 200. Other important skills are SQL, Tableau, R and Python, with all of them being very close to each other in terms of relative importance. Then, after quite a big gap, skills such as PowerBI, AWS and Azure come, clearly seemingly not that meaningful to learn as a marketing analist.

Percentage per skill for Marketing Analist job listings
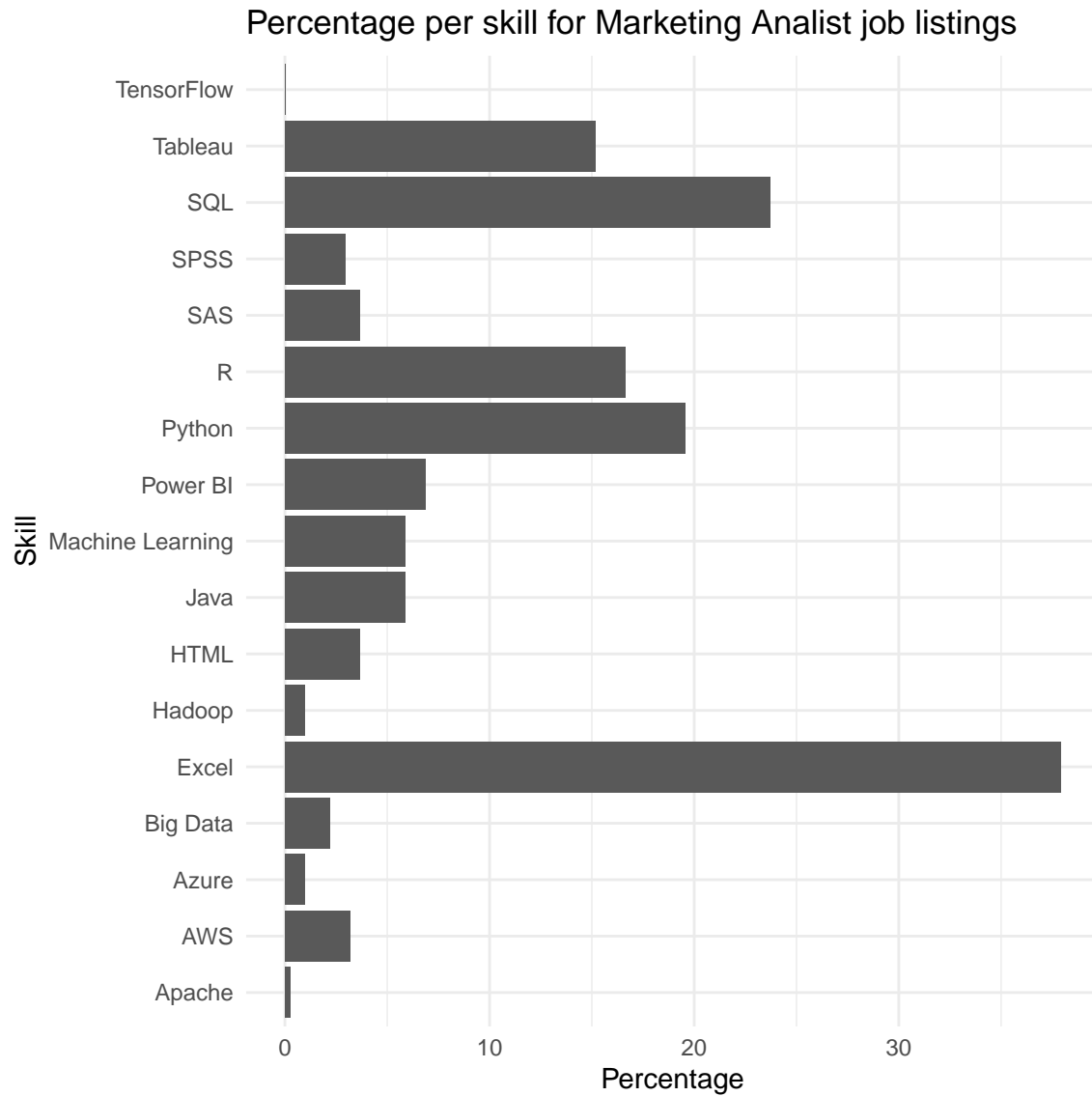
The percentage plot shows very similar results to the count plot. Again, Excel is the clear winner, which is mentioned in over 1/3 of the job descriptions. SQL, Python, R and Tableau follow on respected distance with all being roughly equally relevant for marketing analists to learn.

4.4 Getting the skill counts for the most important skills in marketeer jobs

Counts per skill for Marketeer job listings
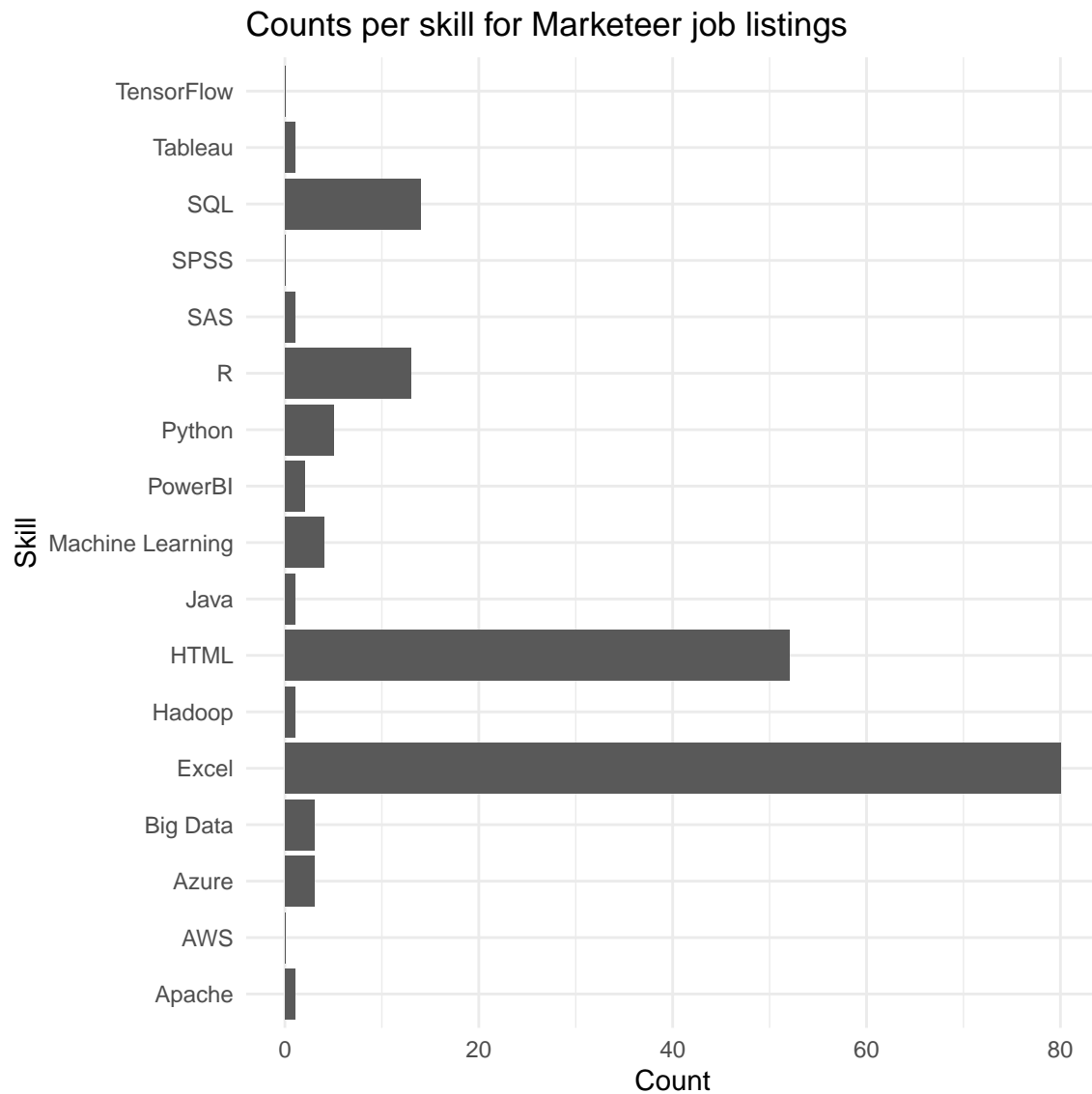
For marketeers, it is clear that technical skills are less important than for the other jobs, which involve more data related tasks. Despite this fact, Excel is still valuable, being the clear winner for marketeers in terms of technical skills. The second spot is for HTML, which is seen as a valuable skill for marketeers to have. Other slightly important skills are R and SQL, with each of the other skills being almost equal to zero in terms of occurrences in the job ads for marketeers.

Percentage per skill for Marketeer job listings
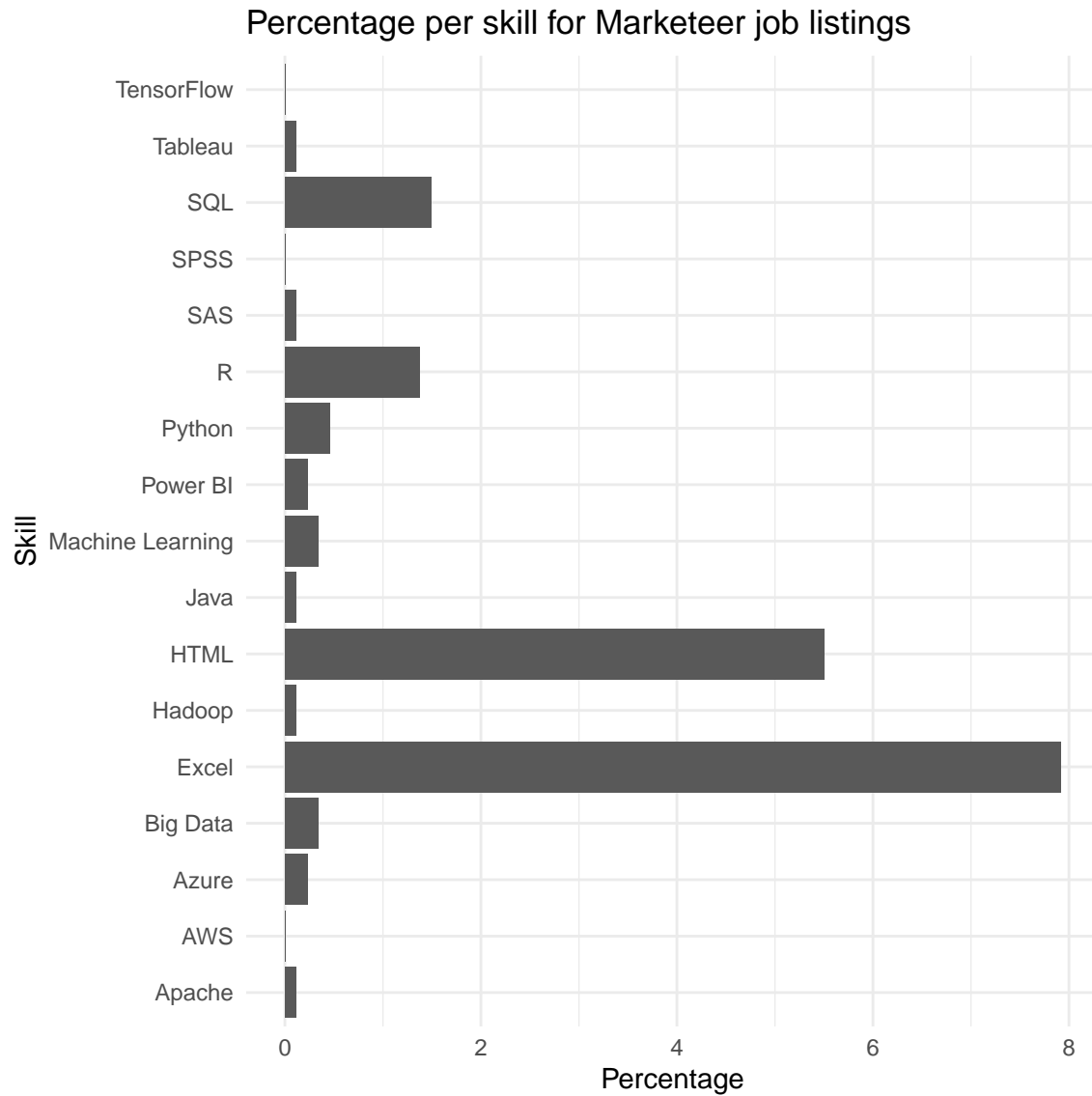
Similar results in percentage, which show that the technical skills get mentioned really infrequently in the job advertisements. Excel stays the winner in terms of technical skills but is still getting mentioned in only eight percent of the job listings. Compared with the winner skills for other job searches, that were reaching over 30-40%, there is a clear difference in the job requirements for marketeers.

4.5 Getting insights into comparing the skills from the 4 job titles

Combining the separate bar charts presented above into one bar graph, to conveniently compare the skills occurrence for our 4 job searches. In the following section we will provide a comprehensive graph which shows each skill percentage occurrence for each job next to each other in order to facilitate direct comparison.

Percentage per skill for all job listings
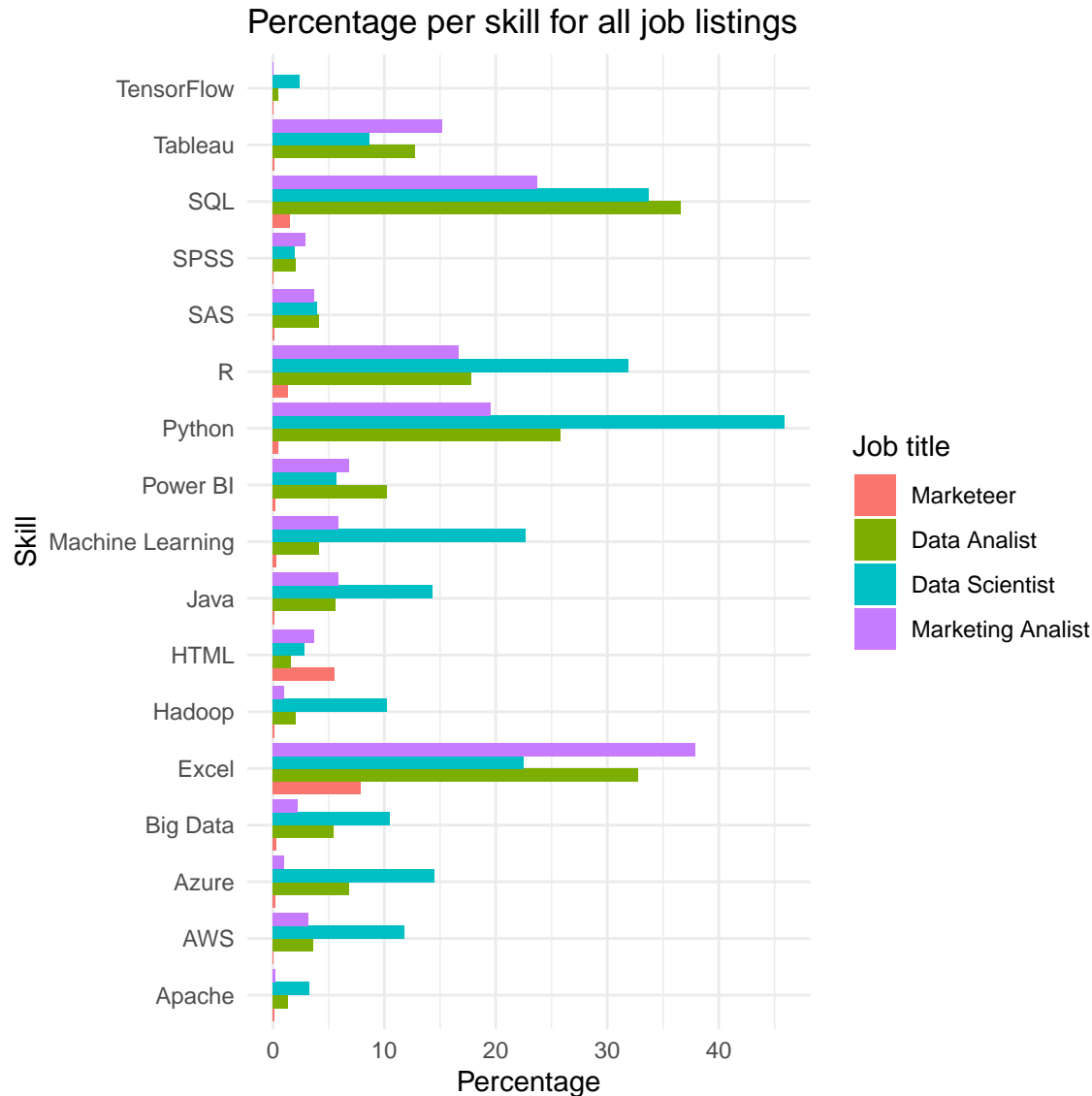
The top 4 skills in general are Python, R, SQL and Excel. With data scientist relying heavily on Python, for marketing and data analists it might be more fruitful to invest in your SQL skills. There seems to be a clear discrepancy between the R and SPSS heavy academic courses and the actual skills required by companies, more focused on SQL and Python to perform job tasks. Especially SQL is completely missing from the MSc Marketing Analytics curriculum. Similarly, Python is available in a few courses only.

5. Location frequency analysis

5.1 Location frequency data science

## Data Science job listings per location
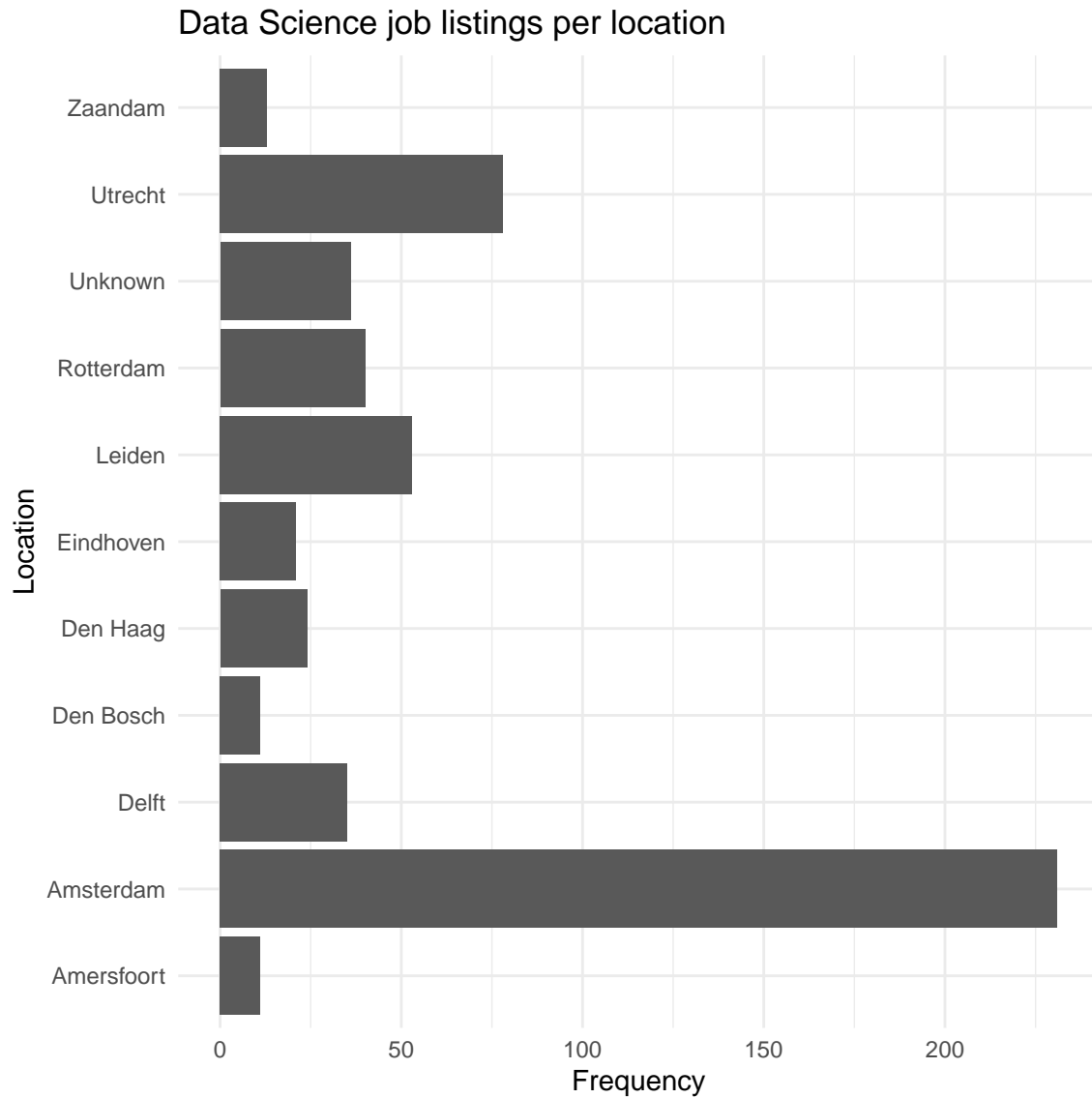
Amsterdam seems the place to be as a data scientist, with counts over 200, or roughly 1/3 of the entire number of job listings posted being in Amsterdam. Furthermore, as expected, the Randstad features the most in job listings, with Utrecht, Rotterdam and Leiden coming in high as well. Outside of the Randstad, Noord-Brabant in the form of Eindhoven and Den Bosch provide a decent number of job opportunities.

Data Analist job listings per location
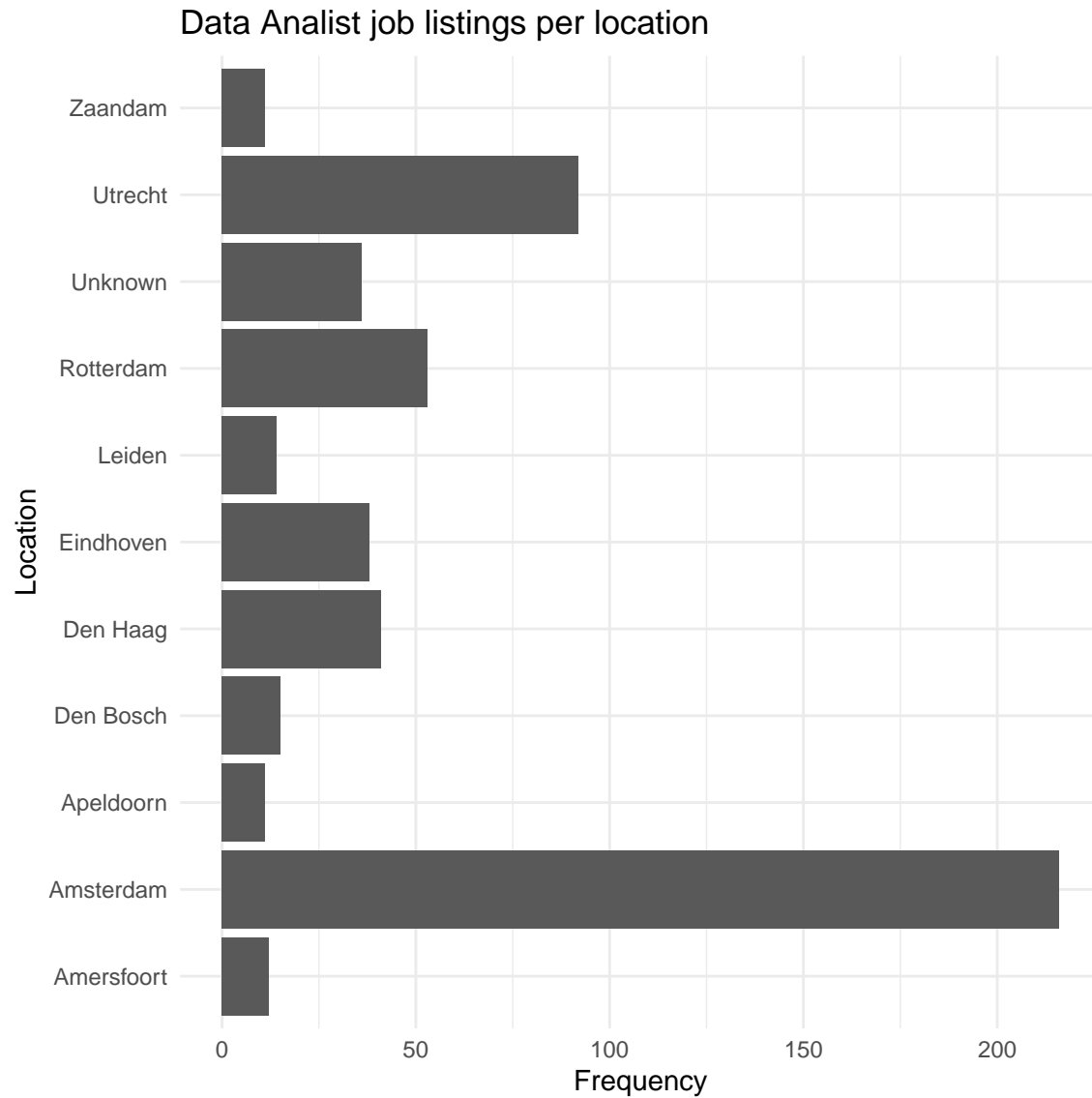
5.2 Location frequency data analist

For data analists, the picture is quite similar. Amsterdam again is the clear winner in terms of number of jobs offered. However, Utrecht seems relatively more important for data analists compared to data scientists. Rotterdam again is number 3, with Den Haag and Eindhoven completing the top 5 in terms of locations.

Marketing Analist job listings per location
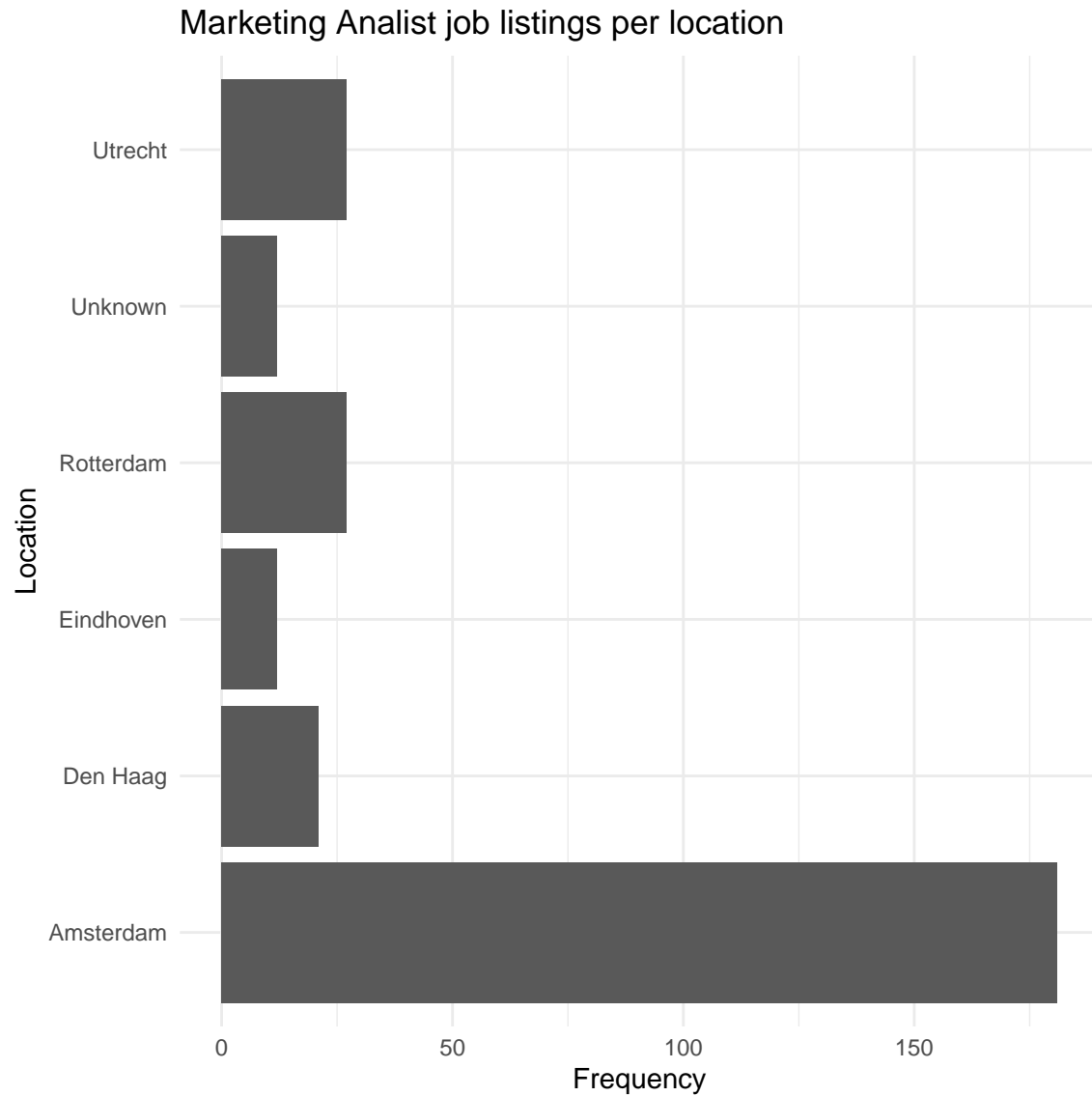
5.3 Location frequency marketing analist

For marketing analists we see the same top 5 as for data analists. However, Utrecht, Rotterdam and Den Haag all have almost equal counts of job offers and Amsterdam is a way more pronounced winner for Marketing analists than for the other jobs scrutinized before.

Marketeer job listings per location
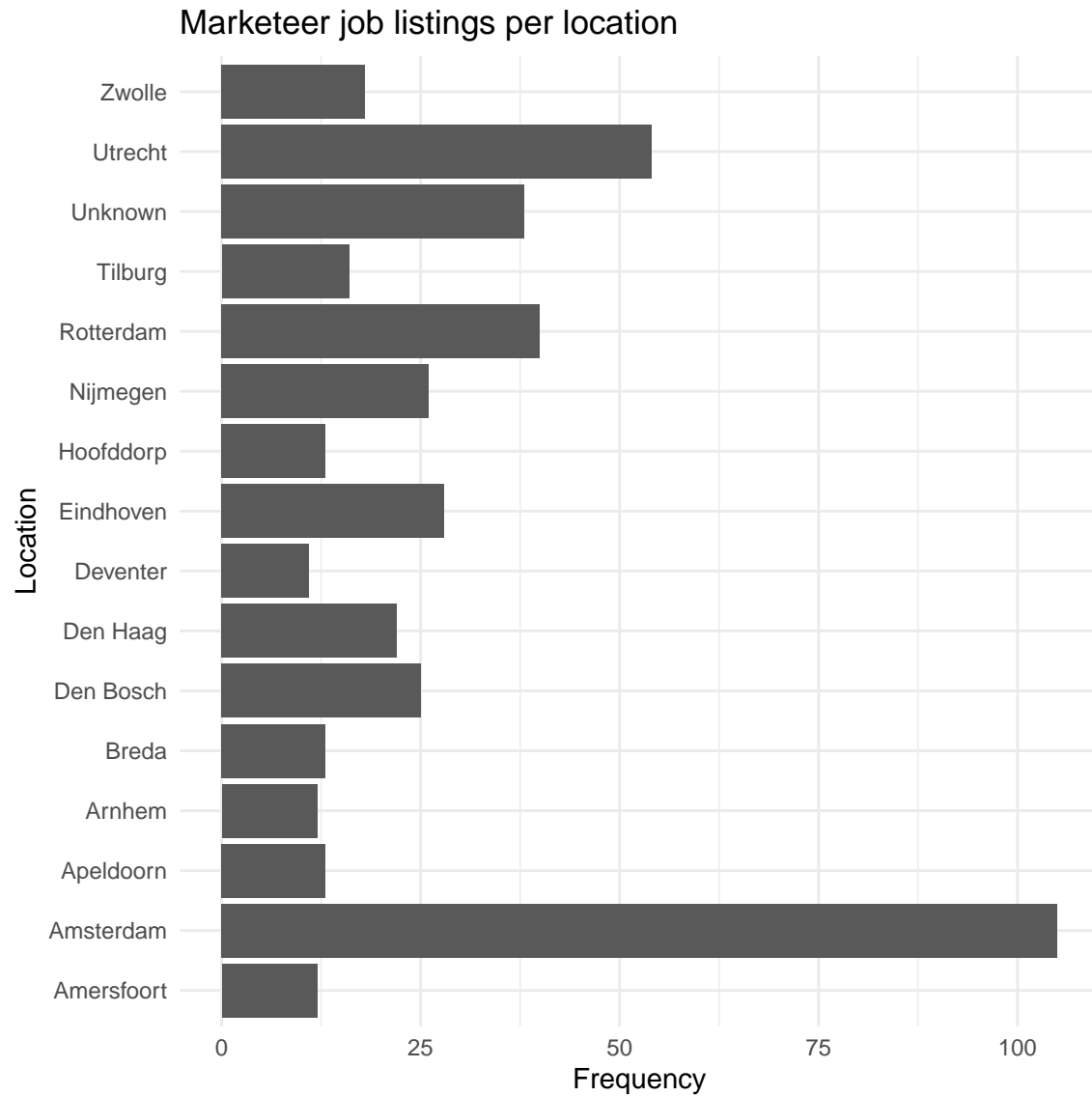
5.4 Location frequency marketeer

For marketeers, it seems that there are more opportunities outside of the Randstad available compared to the three data related jobs. Cities as Breda, Arnhem, Deventer and Nijmegen come in quite close to the Randstad cities such as Rotterdam and Utrecht, suggesting that the job opportunities for marketeers are more evenly spread across the country.

Frequency per location
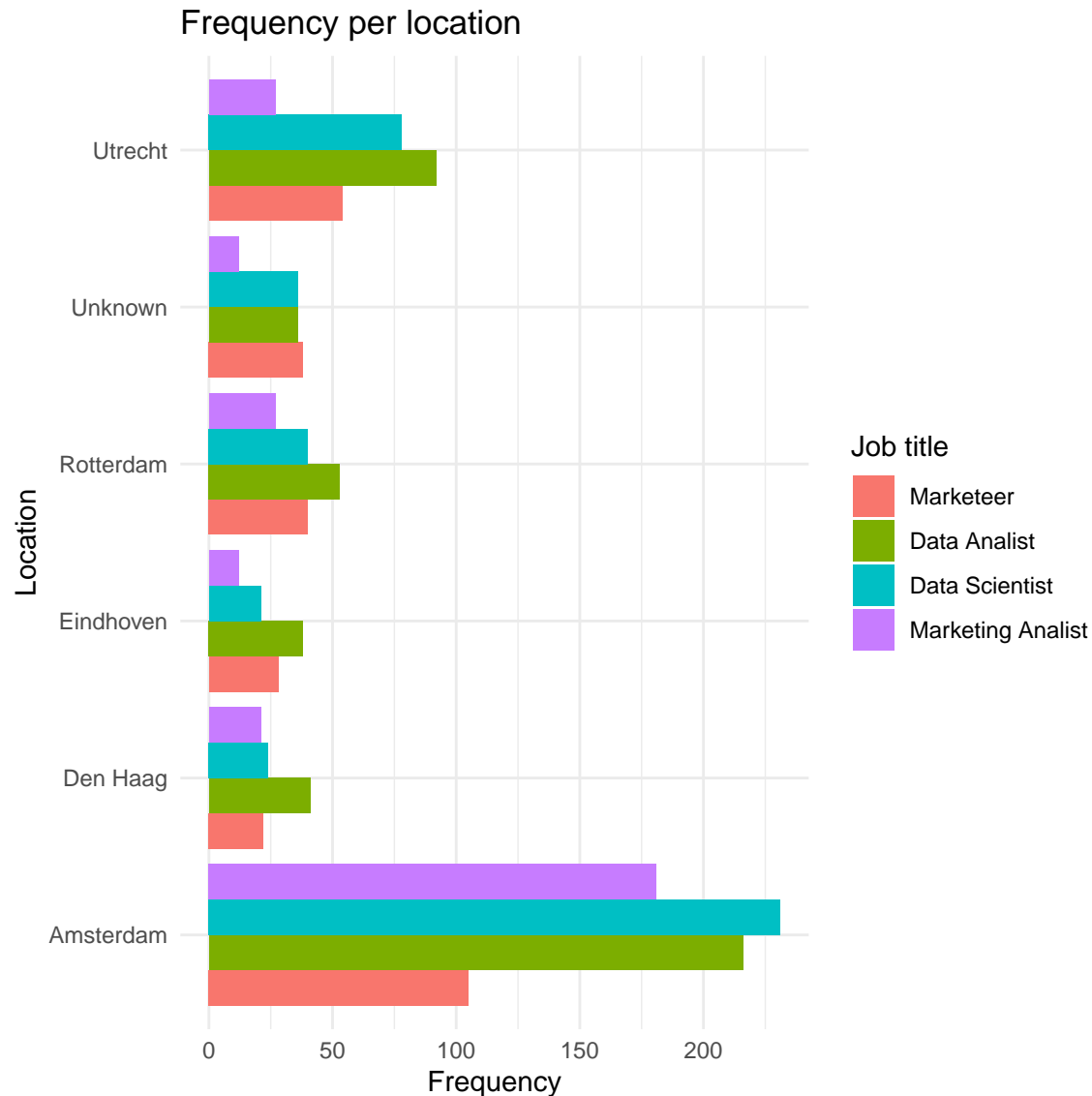
## 5.5 Getting insights into comparing the locations of the 4 job titles

This graph shows the top 5 (plus remote) locations for job listings. As expected, Amsterdam is the clear winner for each job in terms of number of job offers, however the difference is less pronounced for marketeer related jobs than for the other three jobs. Utrecht comes in second in terms of number of job listings whereas Rotterdam, Eindhoven and Den Haag are very equally ranked.

6. Salary analysis

Now that we know what the most important skills are to learn for our four jobs titles and also what the best cities are to go to, we will analyze the best jobs and locations in terms of pay.

6.1 Mean salaries per job

Average salary per month
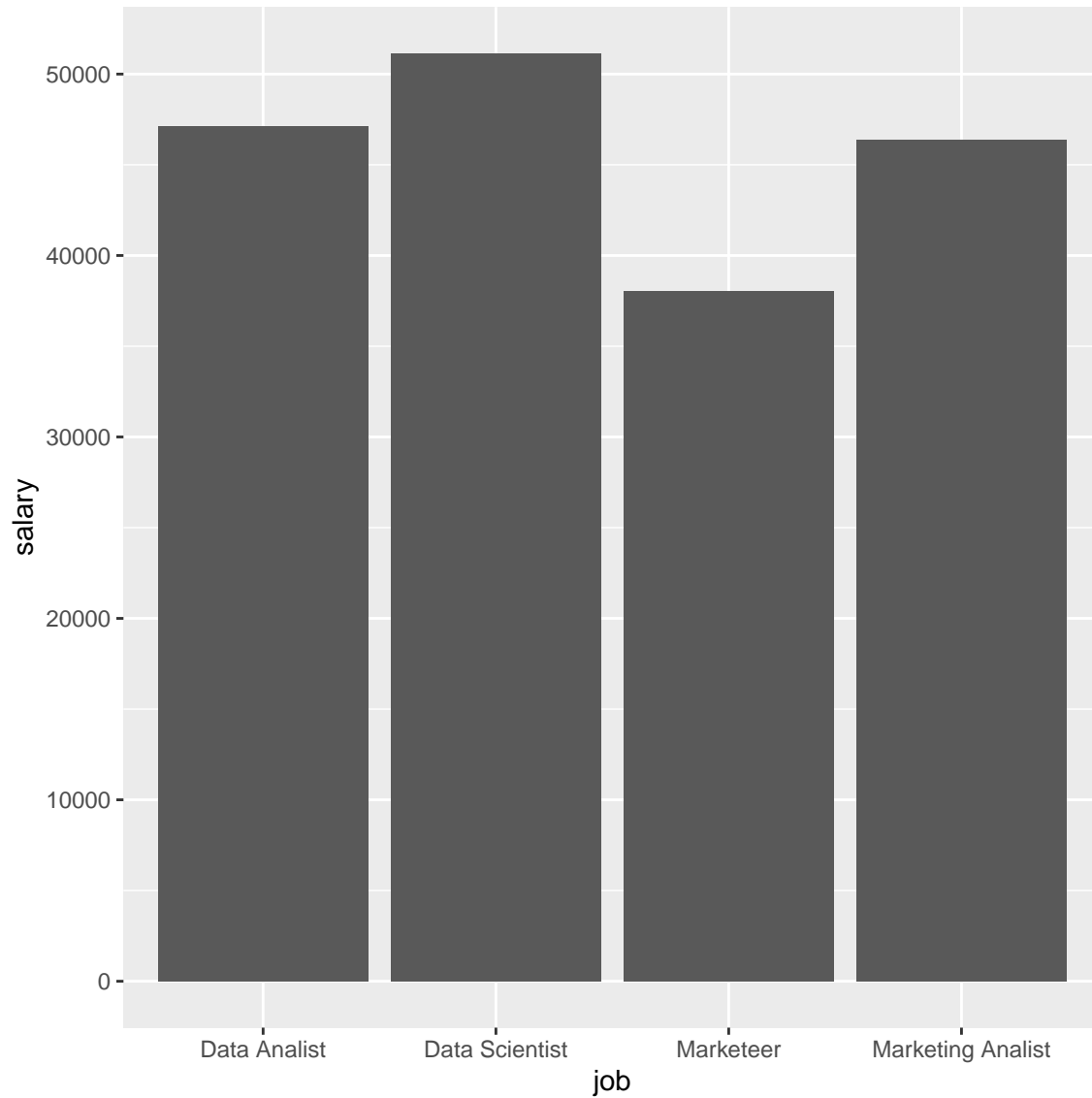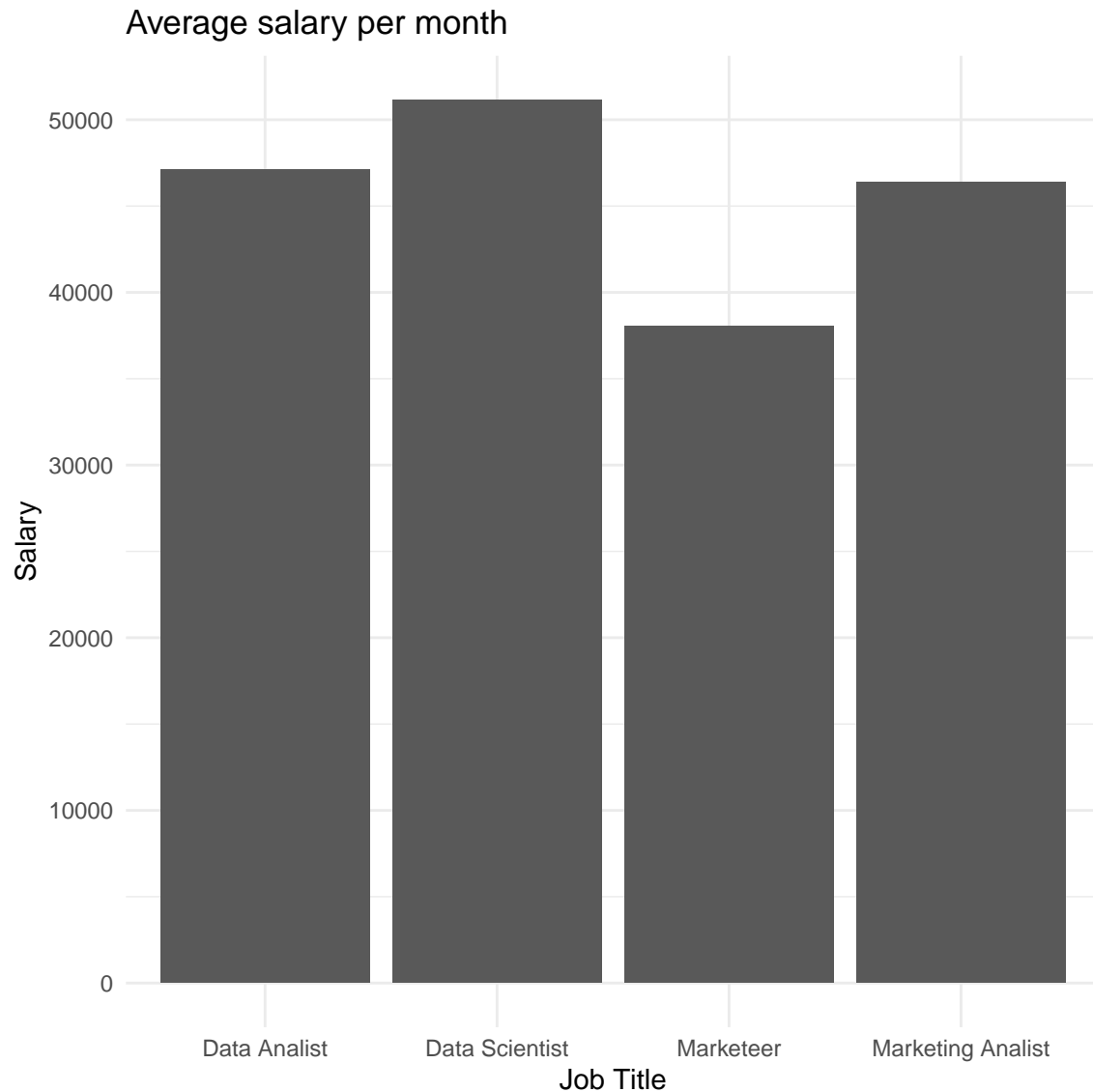
Data Scientists on average have the highest salaries, earning around 50000 annually on average. The difference with Marketing Analysts and Data Analysts is small however, with both analyst jobs having a mean salary of well over 40000 per year. Marketeers seem to fall behind in terms of salary, earning less than 40000 per year.

In the next section we analyse the top locations based on average salaries. Due to limited salary data, many locations have only a single job ad with salary data included, this number does not really represent an average. Therefore, we decided to only incorporate locations with a minimum of 3 job ads into the analysis.

Average salary per location for Data Scientist
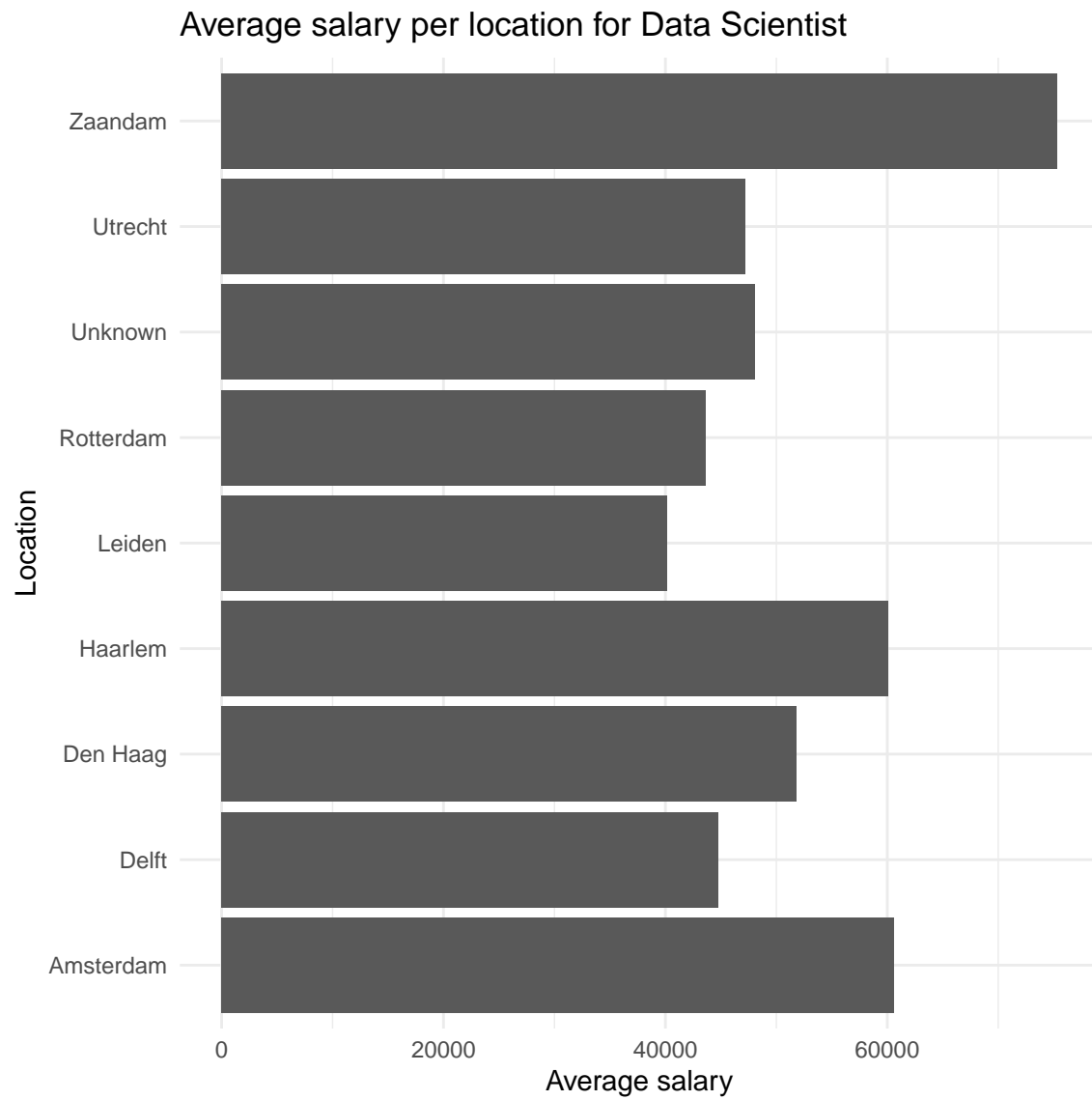
6.2 Top locations in terms of salary for data scientists

## Average salary per location for Data Analists

6.3 Top locations in terms of salary for data analysts

For Data Analysts, again Zaandam and Amsterdam are the cities with highest average salaries. Another attractive option is Apeldoorn. These top 3 cities all have average salaries close to or even over 60000 euro's per year. Quite more than the national average.

Average salary per location for Marketing Analists
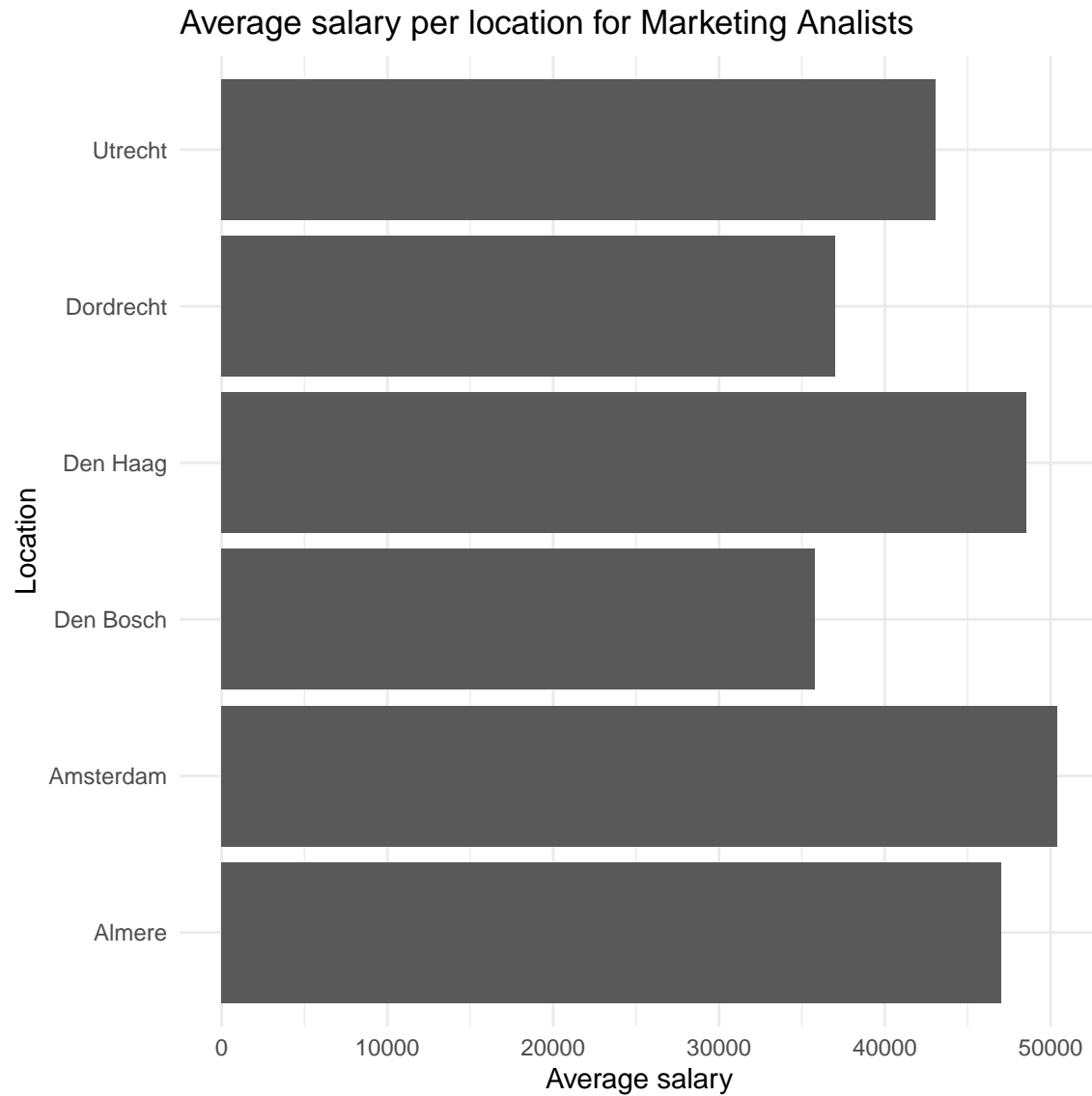
6.4 Top locations in terms of salary for marketing analysts

For Marketing Analysts, when looking for the highest salaries on average, the best places to find a job are Amsterdam, Almere, Den Haag and Utrecht.

Average salary per location for Marketeers
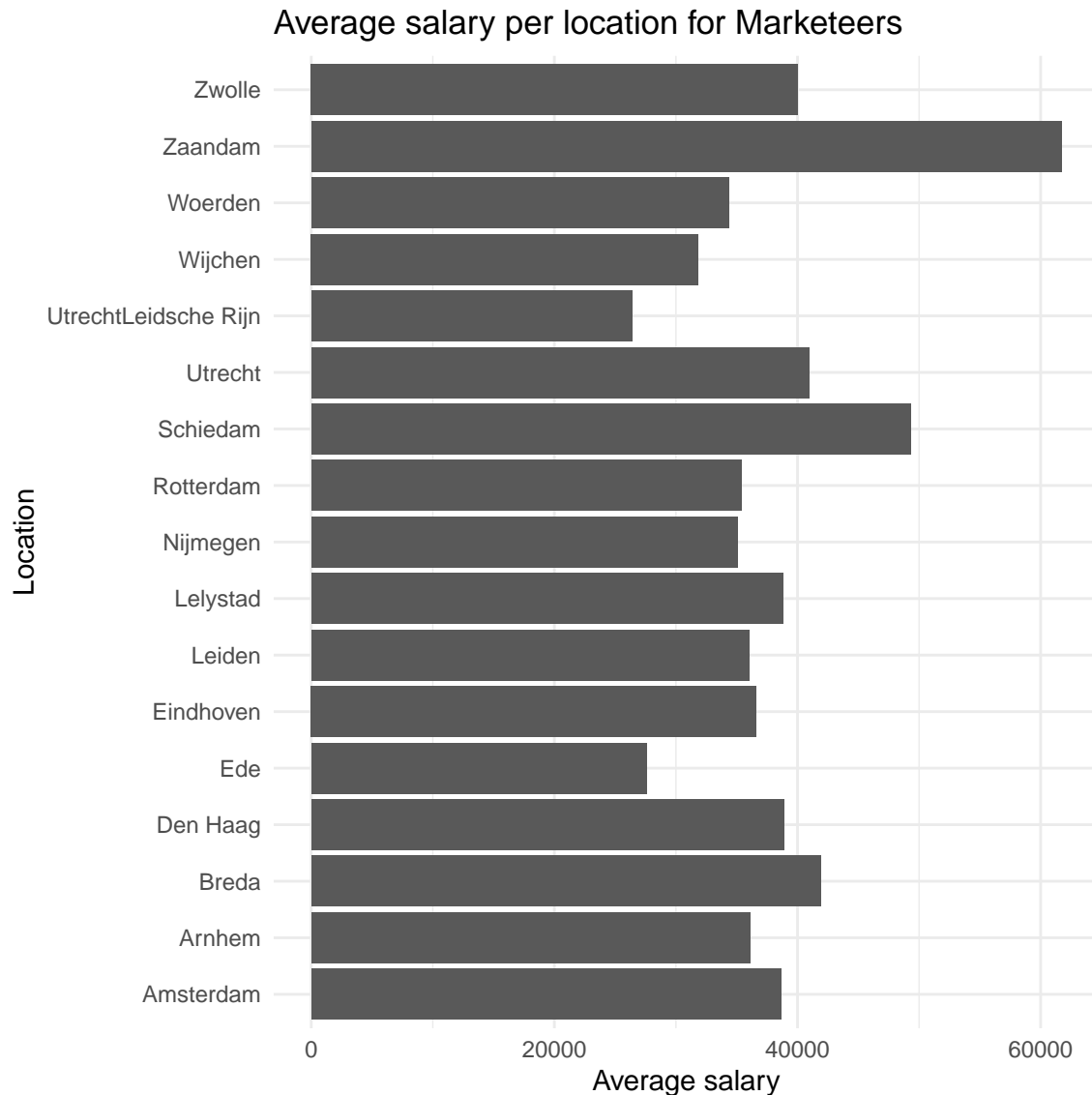
6.5 Top locations in terms of salary for marketeers

The highest paid jobs for marketeers generally are found in Zaandam and Schiedam, with Zaandam having salaries over 60000, which is nearly twice the average salary for marketeers. Besides these 2 cities, there are numerous other cities including Utrecht, Amsterdam, Rotterdam, Nijmegen and Eindhoven all pretty similar to each other in terms of salaries, coming close to 40000 euro average annual income.

Average salary per location
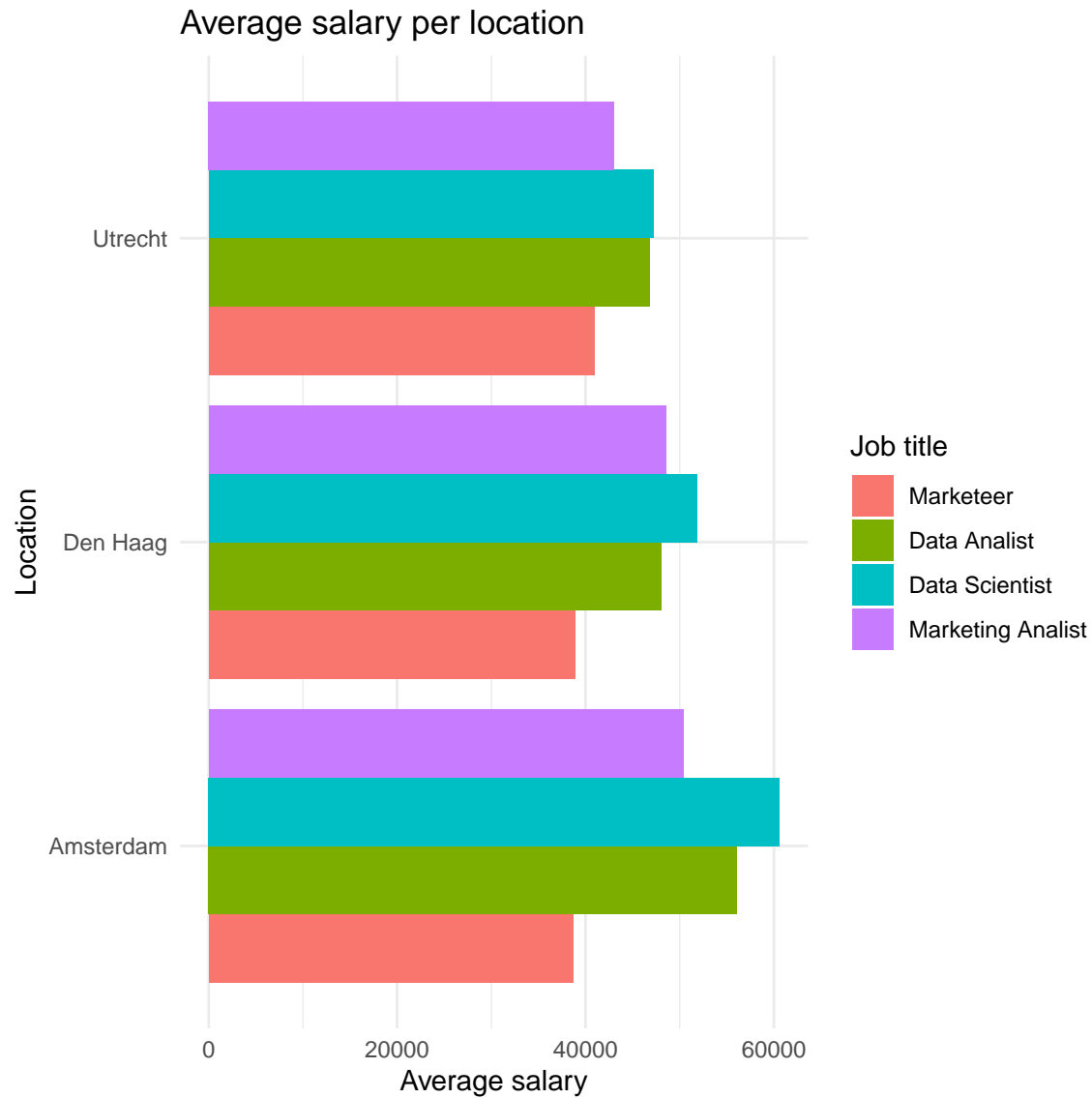
6.6 Getting insights into comparing the salaries of the 4 job titles on the best locations

This plot shows that data scientists in general dominate salaries in each location, shortly followed by data analists and marketing analists. Marketeers in general lag quite behind in terms of salaries.

7 Conclusion and discussion

7.1 Conclusion

Our results show that the main skills required by jobs related to Data Analysis, Data Science and Marketing are Excel, SQL, R and Python. The curriculum of Marketing Analytics is focused more on R and SPSS, and completely ignores SQL. We believe that students would benefit from the inclusion of courses which develop skills in SQL and Python.

As for location, Amsterdam is the clear winner when it comes to number of job offers for any of our 4 search terms. Utrecht comes in second place and Rotterdam, Den Haag and Eindhoven make up the top 5. The results make sense since these are also the biggest cities in terms of population within the Netherlands. The results were pretty similar for the three data jobs but we saw a clear difference for marketeer jobs, where relatively more opportunities were situated outside of the Randstad in cities such as Nijmegen, Breda and Arnhem.

As for salary, data scientists seem to be in the best position, with on average the highest salaries compared to marketeers, data analists and marketing analists.

7.2 Discussion

Our research is limited due to limited sample size. Scraping Indeed.com had its fair share of challenges and our focus on the Netherlands meant that we collected between 500 and 1100 observations per job search. However, after cleaning out duplicates these numbers got adjusted downward. Especially for the salary analysis, there was a lot of missing data which meant datasets with only a few hundred observations. This means that some of the location averages for salary could be very high or very low due to only one job being offered in that location with perhaps a very high or very low salary. This suggests that the salary analysis is sensitive to outlier datapoints. This limited sample is however difficult to overcome, since most job listings do not show salary data.

7.3 Future research

Our results can be reproduced for any job title and any location in the world. This makes it very interesting for future research to compare different countries with each other in terms of salary and job skills for the four jobs researched in this analysis. However, it also opens up the opportunity for students and job seekers in general to perform similar research in any kind of other field they are interested in. By tweaking the search term in the scraper and the skills being asked in the keyword analysis, it is possible to research for example what skills are important in a range of jobs and locations. Furthermore, we think that our analysis could be valuable over time by for example repeating the research on a monthly or yearly basis to observe possible developments over time in skills required, salaries offered and locations offering the most jobs.