

Job market analysys for data roles

Author: *Mateusz Szczepaniak*

Overview

Welcome to my analysis of the data job market, focusing on data analyst roles. This project was created out of a desire to navigate and understand the job market more effectively. It shows the top-paying and in-demand skills to help find optimal job opportunities for data analysts. I explore key questions such as the most demanded skills, salary trends, and the intersection of demand and salary in data analytics.

The Questions

Below are the questions I want to answer in my project:

1. What are the skills most in demand for the top 3 most popular data roles?
2. How are in-demand skills trending for Data Analysts?
3. How well do jobs and skills pay for Data Analysts?
4. What are the optimal skills for data analysts to learn? (High Demand AND High Paying)

Tools I Used

For my deep dive into the data analyst job market, I used tools such as:

- **Python:** allowing me to analyze the data and find critical insights. I also used the following Python libraries:
 - **Pandas Library:** This was used to analyze the data.
 - **Matplotlib Library:** I visualized the data.
 - **Seaborn Library:** Helped me create more advanced visuals.
- **Jupyter Notebooks:** The tool I used to run my Python scripts which let me easily include my notes and analysis.
- **Visual Studio Code:** My go-to for executing my Python scripts.
- **Git & GitHub:** Essential for version control and sharing my Python code and analysis, ensuring collaboration and project tracking.

For my GitHub page go to: github.com/mtszczepaniak

Data Preparation and Cleanup

The data is sourced from Luke Barousse's Python Course which provides a foundation for my analysis, containing detailed information on job titles, salaries, locations, and essential skills.

Composed of **785 000 rows** and **17 columns**.

This section outlines the steps taken to prepare the data for analysis, ensuring accuracy and usability. I start by importing necessary libraries and loading the dataset, followed by initial data cleaning tasks to ensure data quality.

```
# Importing Libraries
import ast
import pandas as pd
from datasets import load_dataset
import matplotlib.pyplot as plt
import seaborn as sns

# Loading Data
dataset = load_dataset('lukebarousse/data_jobs')
df = dataset['train'].to_pandas()

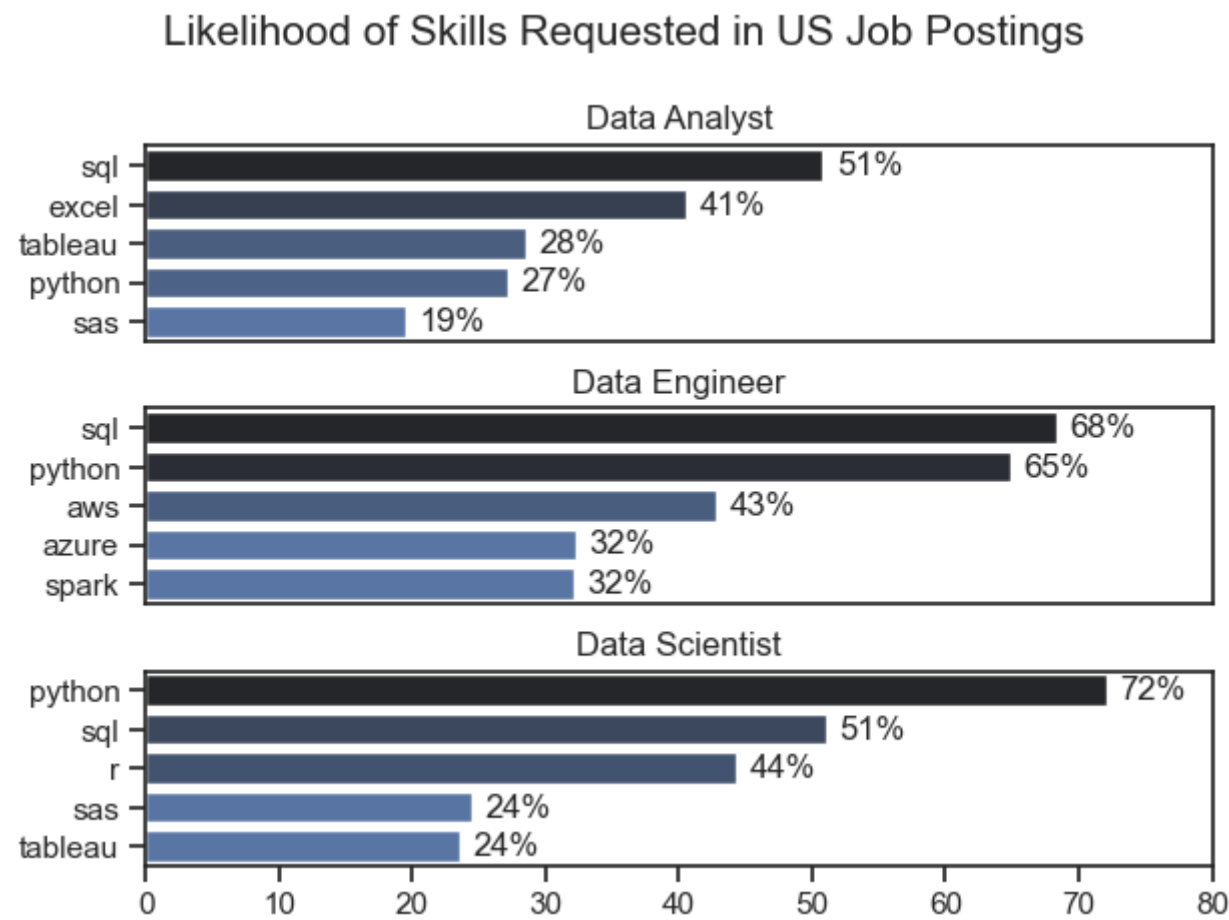
# Data Cleanup
df['job_posted_date'] = pd.to_datetime(df['job_posted_date'])
# Job skills to list
df['job_skills'] = df['job_skills'].apply(lambda x: ast.literal_eval(x) if
pd.notna(x) else x)
```

The Analysis

1. What are the skills most in demand for the top 3 most popular data roles?

To find the most demanded skills for the top 3 most popular data roles. I filtered out those positions by which ones were the most popular, and got the top 5 skills for these top 3 roles. This query highlights the most popular job titles and their top skills, showing which skills I should pay attention to depending on the role I'm targeting.

From all of the three roles selected, those are most often required skills:

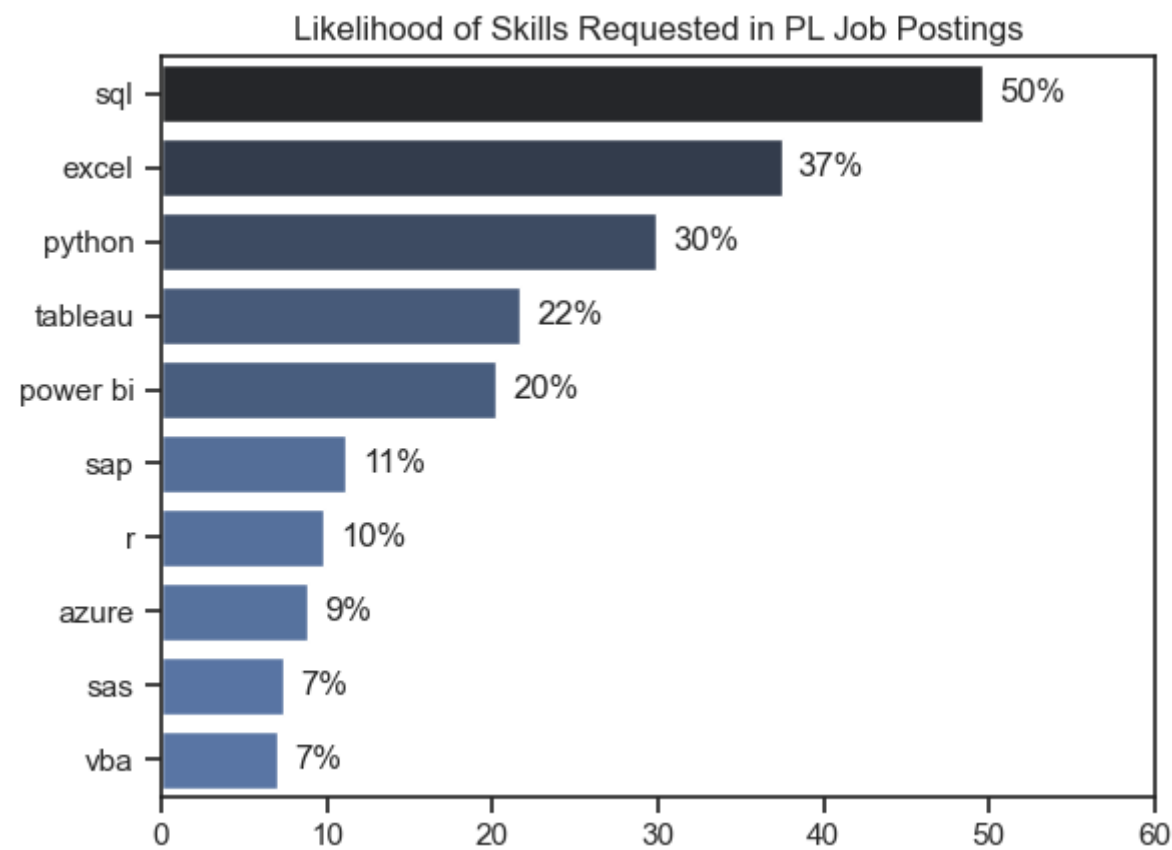


Bar graph visualizing the salary for the top 3 data roles and their top 5 skills associated with each.

Insights:

- SQL is the most requested skill for Data Analysts and Data Scientists, with it in over half the job postings for both roles. For Data Engineers, Python is the most sought-after skill, appearing in 68% of job postings.
- Data Engineers require more specialized technical skills (AWS, Azure, Spark) compared to Data Analysts and Data Scientists who are expected to be proficient in more general data management and analysis tools (Excel, Tableau).
- Python is a versatile skill, highly demanded across all three roles, but most prominently for Data Scientists (72%) and Data Engineers (65%).

For further analysis, I created a data frame containing data from Poland for the role of data analyst. The collected data was used to create the visualization below.

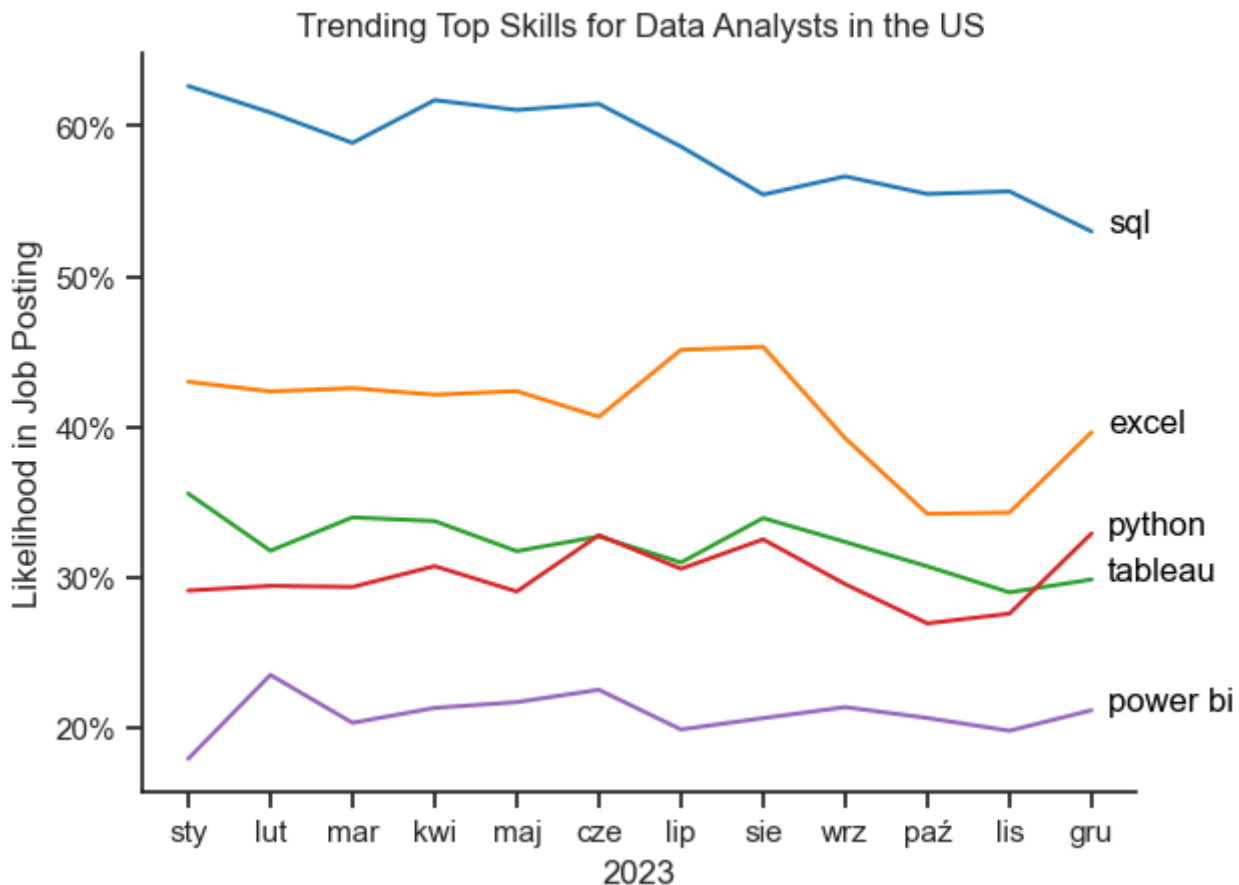


Bar graph visualizing 10 most required skills for Data Analys in Poland.

For data analyst role SQL remains most in demand skill with other popular options behind it such as excel, python, tableau and PowerBi.

2. How are in-demand skills trending for Data Analysts??

To find how skills are trending in 2023 for Data Analysts, I filtered data analyst positions and grouped the skills by the month of the job postings. This got me the top 5 skills of data analysts by month, showing how popular skills were throughout 2023.



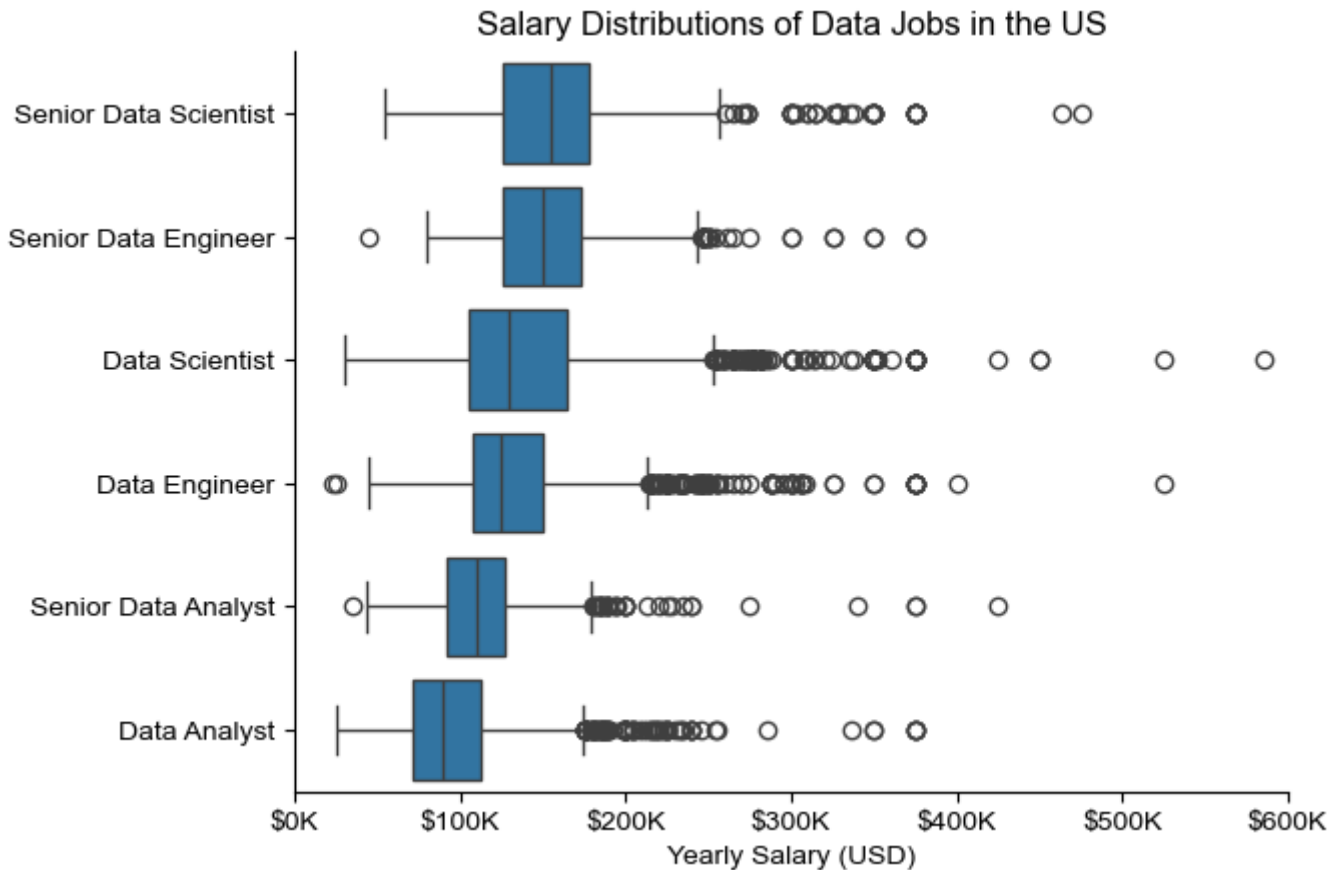
Bar graph visualizing the trending top skills for data analysts in the US in 2023.

Insights:

- SQL remains the most consistently demanded skill throughout the year, although it shows a gradual decrease in demand.
- Excel experienced a significant increase in demand starting around September, surpassing both Python and Tableau by the end of the year.
- Both Python and Tableau show relatively stable demand throughout the year with some fluctuations but remain essential skills for data analysts. Power BI, while less demanded compared to the others, shows a slight upward trend towards the year's end.

3. How well do jobs and skills pay for Data Analysts?

To identify the highest-paying roles and skills, I only got jobs in the United States and looked at their median salary. But first I looked at the salary distributions of common data jobs like Data Scientist, Data Engineer, and Data Analyst, to get an idea of which jobs are paid the most.



Box plot visualizing the salary distributions for the top 6 data job titles.

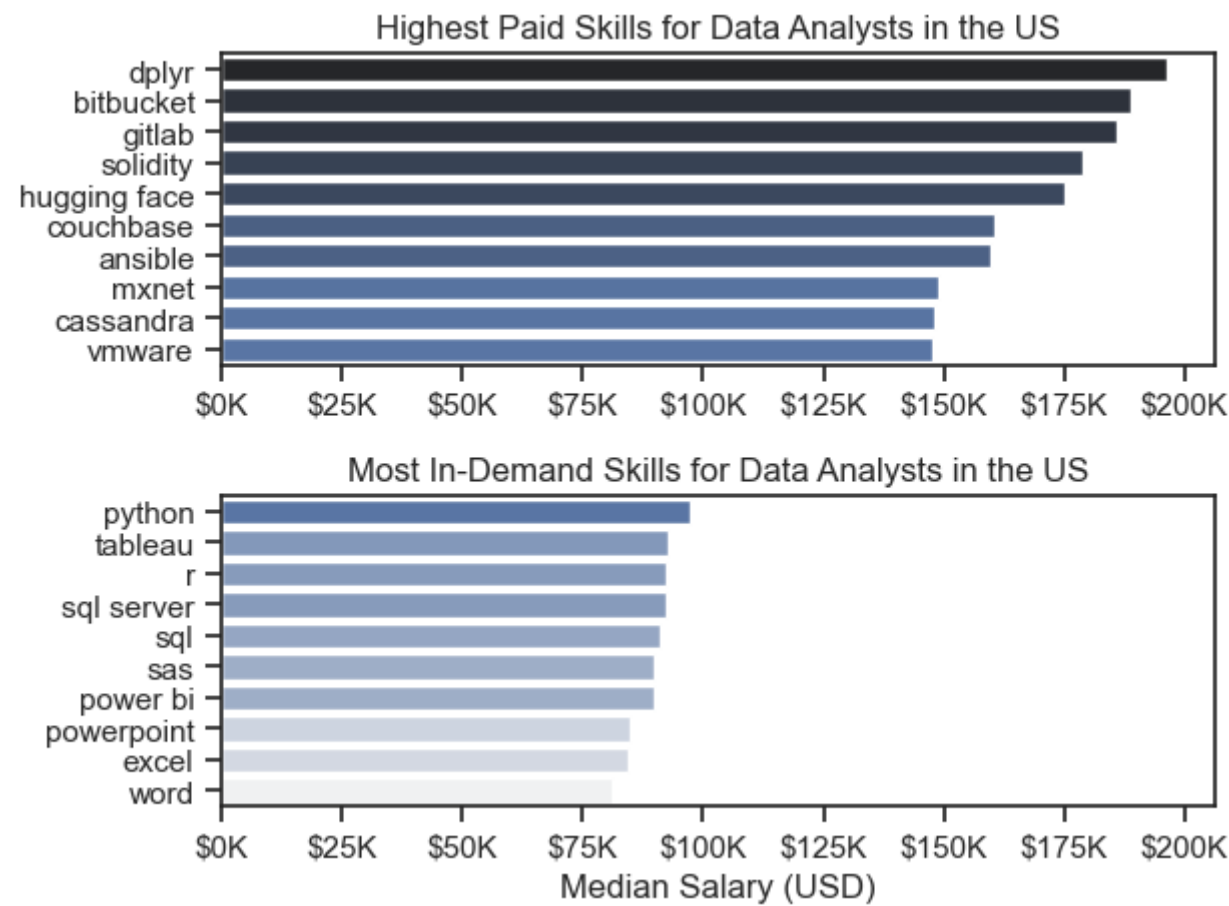
Insights

- There's a significant variation in salary ranges across different job titles. Senior Data Scientist positions tend to have the highest salary potential, with up to \$600K, indicating the high value placed on advanced data skills and experience in the industry.
- Senior Data Engineer and Senior Data Scientist roles show a considerable number of outliers on the higher end of the salary spectrum, suggesting that exceptional skills or circumstances can lead to high pay in these roles. In contrast, Data Analyst roles demonstrate more consistency in salary, with fewer outliers.
- The median salaries increase with the seniority and specialization of the roles. Senior roles (Senior Data Scientist, Senior Data Engineer) not only have higher median salaries but also larger differences in typical salaries, reflecting greater variance in compensation as responsibilities increase.

Highest Paid & Most Demanded Skills for Data Analysts

Next, I narrowed my analysis and focused only on data analyst roles. I looked at the highest-paid skills and the most in-demand skills. I used two bar charts to showcase these.

Here's the breakdown of the highest-paid & most in-demand skills for data analysts in the US:



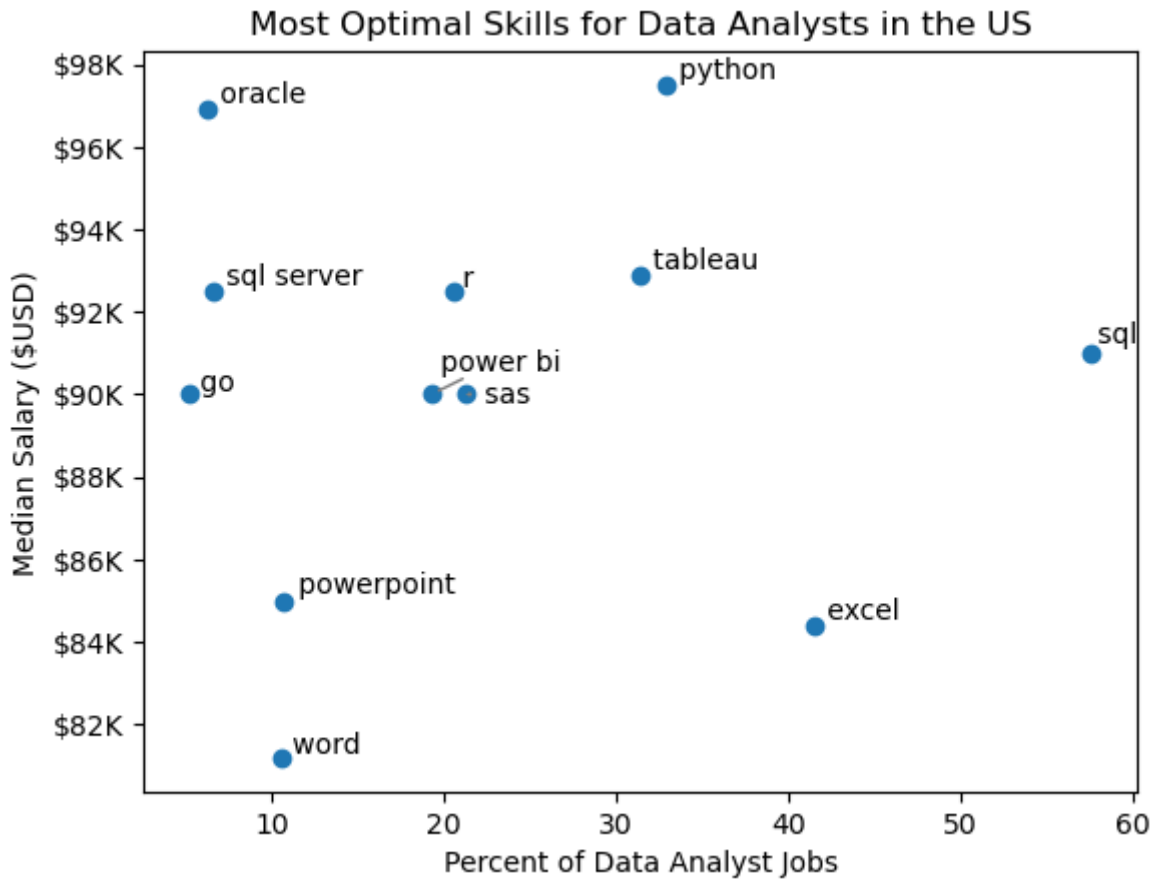
Two separate bar graphs visualizing the highest paid skills and most in-demand skills for data analysts in the US.

Insights:

- The top graph shows specialized technical skills like **dplyr**, **Bitbucket**, and **Gitlab** are associated with higher salaries, some reaching up to \$200K, suggesting that advanced technical proficiency can increase earning potential.
- The bottom graph highlights that foundational skills like **Excel**, **PowerPoint**, and **SQL** are the most in-demand, even though they may not offer the highest salaries. This demonstrates the importance of these core skills for employability in data analysis roles.
- There's a clear distinction between the skills that are highest paid and those that are most in-demand. Data analysts aiming to maximize their career potential should consider developing a diverse skill set that includes both high-paying specialized skills and widely demanded foundational skills.

4. What are the most optimal skills to learn for Data Analysts?

To identify the most optimal skills to learn (the ones that are the highest paid and highest in demand) I calculated the percent of skill demand and the median salary of these skills. To easily identify which are the most optimal skills to learn.



A scatter plot visualizing the most optimal skills (high paying & high demand) for data analysts in the US.

Insights:

- The skill **Oracle** appears to have the highest median salary of nearly \$97K, despite being less common in job postings. This suggests a high value placed on specialized database skills within the data analyst profession.
- More commonly required skills like **Excel** and **SQL** have a large presence in job listings but lower median salaries compared to specialized skills like **Python** and **Tableau**, which not only have higher salaries but are also moderately prevalent in job listings.
- Skills such as **Python**, **Tableau**, and **SQL Server** are towards the higher end of the salary spectrum while also being fairly common in job listings, indicating that proficiency in these tools can lead to good opportunities in data analytics.
- The scatter plot shows that most of the **programming** skills tend to cluster at higher salary levels compared to other categories, indicating that programming expertise might offer greater salary benefits within the data analytics field.

- The database skills, such as **Oracle** and **SQL Server**, are associated with some of the highest salaries among data analyst tools. This indicates a significant demand and valuation for data management and manipulation expertise in the industry.
- Analyst tools, including **Tableau** and **Power BI**, are prevalent in job postings and offer competitive salaries, showing that visualization and data analysis software are crucial for current data roles. This category not only has good salaries but is also versatile across different types of data tasks.

Insights

This project provided several general insights into the data job market for analysts:

- **Skill Demand and Salary Correlation:** There is a clear correlation between the demand for specific skills and the salaries these skills command. Advanced and specialized skills like Python and Oracle often lead to higher salaries.
- **Market Trends:** There are changing trends in skill demand, highlighting the dynamic nature of the data job market. Keeping up with these trends is essential for career growth in data analytics.
- **Economic Value of Skills:** Understanding which skills are both in-demand and well-compensated can guide data analysts in prioritizing learning to maximize their economic returns.

What I Learned

Throughout this project, I deepened my understanding of the data analyst job market and enhanced my technical skills in Python, especially in data manipulation and visualization. Here are a few specific things I learned:

- **Advanced Python Usage:** Utilizing libraries such as Pandas for data manipulation, Seaborn and Matplotlib for data visualization, and other libraries helped me perform complex data analysis tasks more efficiently.
- **Data Cleaning Importance:** I learned that thorough data cleaning and preparation are crucial before any analysis can be conducted, ensuring the accuracy of insights derived from the data.
- **Strategic Skill Analysis:** The project emphasized the importance of aligning one's skills with market demand. Understanding the relationship between skill demand, salary, and job availability allows for more strategic career planning in the tech industry.

Challenges I Faced

This project was not without its challenges, but it provided good learning opportunities:

- **Data Inconsistencies:** Handling missing or inconsistent data entries requires careful consideration and thorough data-cleaning techniques to ensure the integrity of the analysis.
- **Complex Data Visualization:** Designing effective visual representations of complex datasets was challenging but critical for conveying insights clearly and compellingly.
- **Balancing Breadth and Depth:** Deciding how deeply to dive into each analysis while maintaining a broad overview of the data.