

Data Analyst Skills & Salary Analysis with SQL

Introduction

This project explores the intersection of skill demand and salary across job postings, focusing on Data Analyst roles. By leveraging SQL to analyze structured job and skill data, we aim to uncover actionable insights for job seekers, educators, and recruiters.

Key questions we address:

- Which Data Analyst roles offer the highest salaries ?
- What skills are most common in high-paying Data Analyst jobs ?
- Which skills are most in-demand for Data Analyst roles ?
- What are the skills that pay the highest ?
- What are the best skills to learn based on a combination of demand and salary ?

The findings help in prioritizing career development and tailoring educational programs to align with current market needs.

Tools I Used

- **Excel** – Simple visualization and Data Cleaning
 - **PostgreSQL** – SQL querying and aggregation
 - **SQL (PostgreSQL dialect)** – CTEs, joins, filters, math functions
 - **Git & GitHub** – Version control and collaboration
 - **VS Code** – Code writing and editing
-

Analysis

1. Highest-Paying Data Analyst Roles

Problem Statement:

Identify the highest-paying Data Analyst job postings to help job seekers target high-value opportunities.

Description:

This query filters out job postings without salary information, converts hourly wages to annual if necessary, and returns the top 10 roles offering the highest pay.

Approach:

- Filter for valid `salary_year_avg` or `salary_hour_avg`.
- Convert hourly salaries to annual equivalents using **2040** hours/year.
- Filter job title as "Data Analyst".
- Join with the `company_dim` table for company names.
- Order by salary in descending order.

```

SELECT
    job_id,
    job_title,
    COALESCE(salary_year_avg, salary_hour_avg*2040) AS salary,
    name AS company_name,
    job_location,
    job_schedule_type
FROM
    job_postings_fact
LEFT JOIN
    company_dim ON job_postings_fact.company_id = company_dim.company_id
WHERE
    (salary_year_avg IS NOT NULL OR salary_hour_avg IS NOT NULL)
    AND job_title_short = 'Data Analyst'
ORDER BY
    salary DESC
LIMIT 10;

```

2. Most In-Demand Skills for High-Paying Data Analyst Roles

Problem Statement:

Identify the most common skills associated with the highest-paying Data Analyst job postings.

Description:

This analysis focuses on uncovering the skills most frequently associated with top-paying Data Analyst roles. By selecting the 100 highest-paying job postings for Data Analysts (based on annualized salary), we analyze the skills linked to these positions. The result reveals the 10 most common skills among these high-salary jobs — offering insights into the competencies that tend to appear in well-compensated Data Analyst roles.

Approach:

- Filtered the job postings dataset to include only Data Analyst roles.
- Calculated a unified salary field by converting hourly wages to annual salaries where needed.
- Selected the top 100 jobs with the highest salaries.
- Joined these jobs with the skills dataset to retrieve associated skills.
- Grouped and counted the number of job postings each skill appeared in.
- Ordered the skills by frequency and selected the top 10 most common skills.

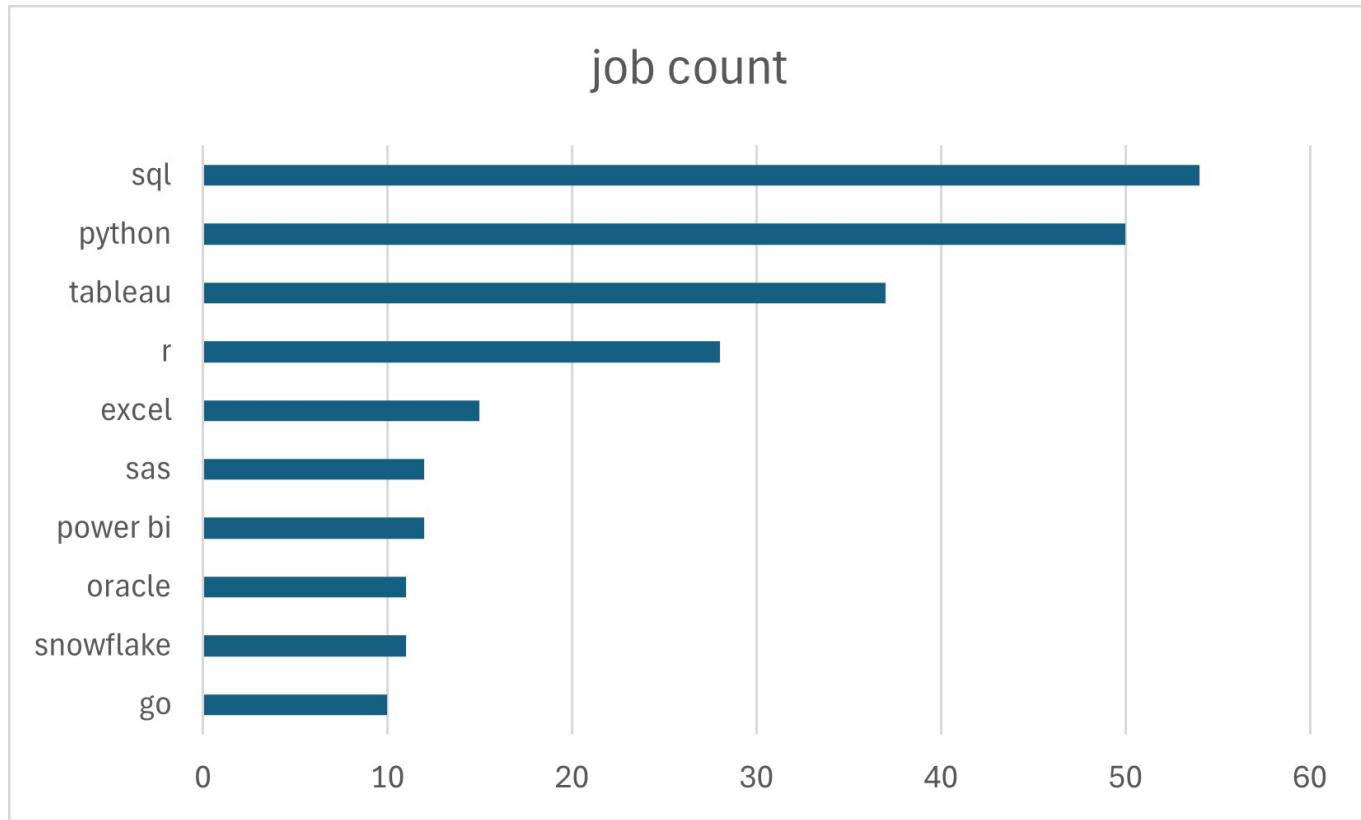
```

WITH job_skills AS(
    SELECT
        job_id,
        job_title,
        COALESCE(salary_year_avg, salary_hour_avg*2040) AS salary
    FROM
        job_postings_fact
    WHERE
        (salary_year_avg is NOT NULL OR salary_hour_avg is NOT NULL)
        AND job_title_short = 'Data Analyst'
    ORDER BY

```

```
    salary DESC
LIMIT
100
)

SELECT
skills_dim.skills,
count(*) as "job count"
FROM
job_skills
INNER JOIN
skills_job_dim on skills_job_dim.job_id = job_skills.job_id
INNER JOIN
skills_dim on skills_job_dim.skill_id = skills_dim.skill_id
GROUP BY
skills_dim.skills
ORDER BY
"job count" DESC
LIMIT
10;
```



This chart highlights the skills most frequently associated with the top 100 highest-paying Data Analyst roles, offering insight into the technical competencies valued in premium positions.

3. Most Frequently Required Skills for Data Analyst Roles

Problem Statement:

Identify the top most frequently required skills for Data Analyst job roles.

Description:

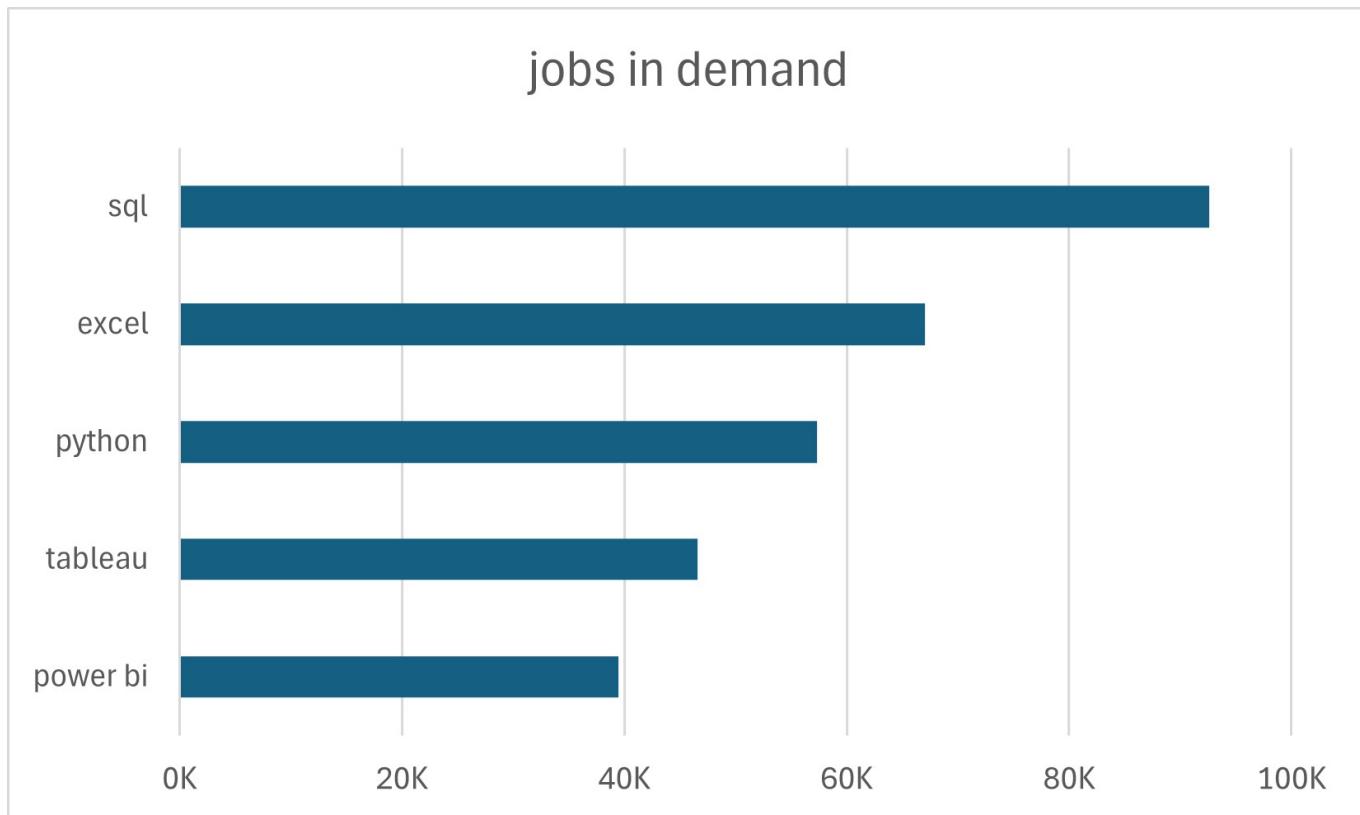
In today's competitive job market, knowing the most sought-after skills can significantly improve job-seeking strategies and curriculum development. This query aims to uncover the top 5 most in-demand skills specifically for Data Analyst roles, based on job postings.

The result helps job seekers, hiring managers, and educators understand which skills are most valued in the industry for Data Analyst positions.

Approach:

- Filter job postings to include only those with the title Data Analyst.
- Join the job postings with the associated skills using `skills_job_dim`.
- Count how many job postings mention each skill.
- Order the results by the number of mentions in descending order.
- Limit the output to the top 5 most commonly required skills.

```
SELECT
    skills,
    count.skills_job_dim.job_id as "jobs in demand"
FROM
    skills_job_dim
INNER JOIN
    job_postings_fact ON job_postings_fact.job_id = skills_job_dim.job_id
INNER JOIN
    skills_dim ON skills_dim.skill_id = skills_job_dim.skill_id
WHERE
    job_postings_fact.job_title_short = 'Data Analyst'
GROUP BY
    skills
ORDER BY
    "jobs in demand" DESC
LIMIT 5;
```



This chart shows the number of job postings that mention each skill among the top 100 highest-paying Data Analyst roles, highlighting the most in-demand competencies in premium positions.

4. Highest-Paying Skills Across Job Postings

Problem Statement:

Identify the highest-paying skills across all job postings with valid salary data.

Description:

By ranking skills according to the average normalized salary of the jobs requiring them, this query pinpoints which competencies command the highest pay in the market. Job seekers can prioritize developing these skills for maximum earning potential, and educators can tailor curricula to high-value areas.

Approach:

- Filter job postings to include only those with a valid annual or hourly salary.
- Normalize salaries by converting hourly wages to annual equivalents (assuming 2,040 work hours per year).
- Pre-order the normalized jobs by salary in descending order to optimize for top salaries.
- Join the filtered job list with `skills_job_dim` and `skills_dim` to associate each job with its skills.
- For each skill, compute the average salary of its associated jobs.
- Order skills by average salary in descending order.
- Return the top 25 skills with the highest average salary.

```
WITH job_fact_new AS(
  SELECT
    job_id,
    job_title,
    COALESCE(salary_year_avg, salary_hour_avg*2040) AS salary
```

```

FROM
    job_postings_fact
WHERE
    (salary_year_avg is NOT NULL OR salary_hour_avg is NOT NULL)
    ORDER BY
        salary DESC
)

SELECT
    skills,
    round(avg(salary)::numeric,0) as "average salary"
FROM
    job_fact_new
INNER JOIN
    skills_job_dim on skills_job_dim.job_id = job_fact_new.job_id
INNER JOIN
    skills_dim on skills_job_dim.skill_id = skills_dim.skill_id
GROUP BY
    skills
ORDER BY
    "average salary" DESC
LIMIT
    25;

```

5. Optimal In-Demand Skills Based on Demand and Salary

Problem Statement:

Determine the optimal in-demand skills by analyzing their job market demand alongside the associated average salaries to guide career and hiring decisions.

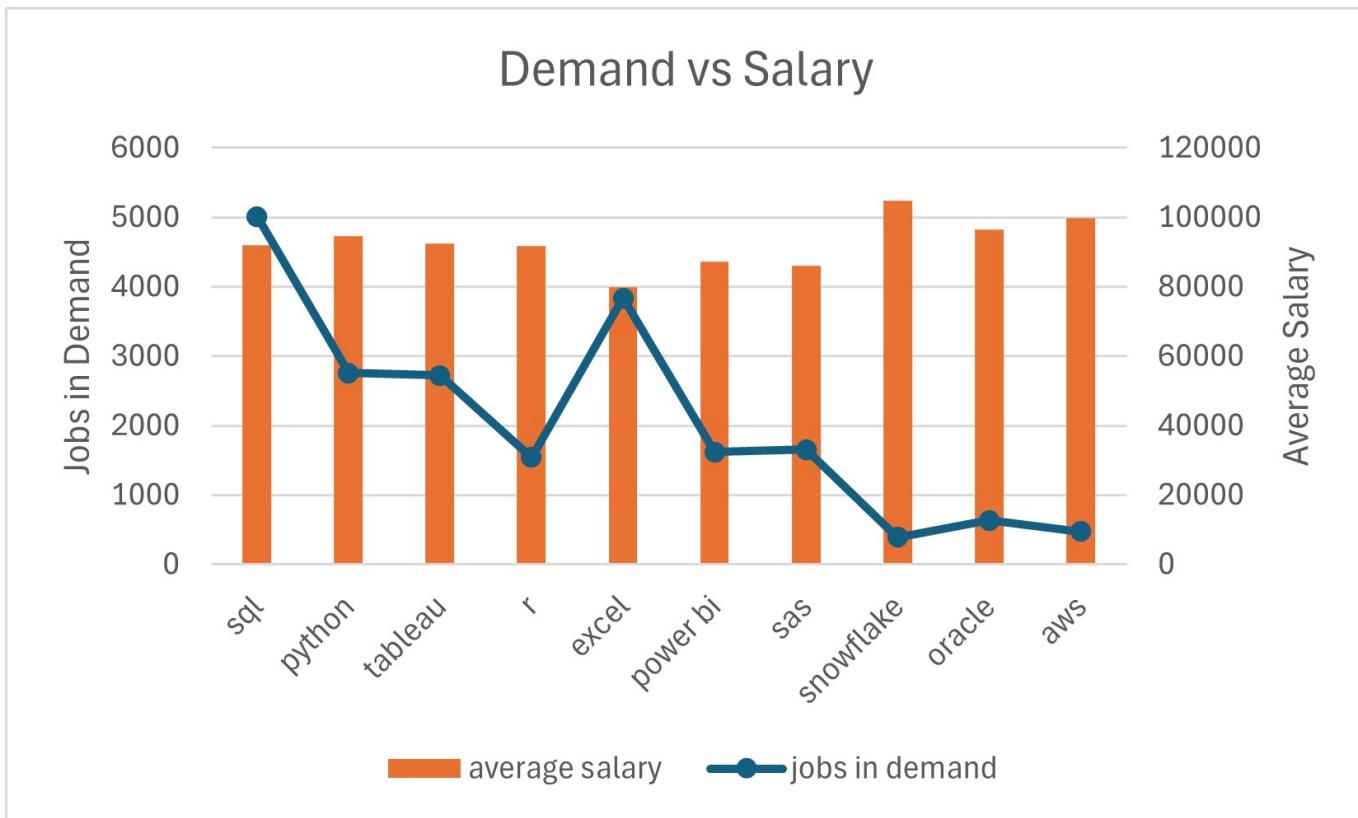
Description:

In the evolving job landscape, job seekers and educators benefit from knowing not just which skills are frequently required but also which offer strong earning potential. This analysis ranks skills by optimizing both their demand (number of job postings) and their average salary, giving a holistic view of valuable skills in the job market.

Approach:

- Filter job postings to include only those with a valid annual or hourly salary.
- Normalize salaries by converting hourly wages to annual equivalents.
- Join job postings with associated skills via `skills_job_dim` and `skills_dim`.
- For each skill, calculate:
 - The number of job postings requiring that skill.
 - The average salary for jobs that require that skill.
- Sort the skills based on a combined score using:
 - **average salary × log(number of jobs)** This scoring formula balances salary potential with market demand, ensuring that highly ranked skills are not just well-paid but also widely required. The logarithmic factor dampens the effect of extremely high job counts, allowing skills with fewer but better-paying jobs to remain competitive in the ranking.
- Return the top 10 skills based on this sorting.

```
WITH job_fact_new AS (
    SELECT
        job_id,
        job_title,
        COALESCE(salary_year_avg, salary_hour_avg * 2040) AS salary,
        job_location,
        job_schedule_type
    FROM
        job_postings_fact
    WHERE
        (salary_year_avg IS NOT NULL OR salary_hour_avg IS NOT NULL)
        and job_title_short = 'Data Analyst'
),
average_count AS (
    SELECT
        skills,
        COUNT.skills_job_dim.job_id AS "jobs in demand",
        ROUND(AVG(salary), 0) AS "average salary"
    FROM
        skills_job_dim
    INNER JOIN
        job_fact_new ON job_fact_new.job_id = skills_job_dim.job_id
    INNER JOIN
        skills_dim ON skills_dim.skill_id = skills_job_dim.skill_id
    GROUP BY
        skills
)
SELECT
    skills,
    "jobs in demand",
    "average salary"
FROM
    average_count
ORDER BY
    ROUND("average salary" * LOG("jobs in demand")::numeric, 0) DESC
LIMIT
    10;
```



This chart presents a dual view of the job market demand (number of job postings) and earning potential (average salary) for key Data Analyst skills. By applying a combined score of **average salary $\times \log(\text{job demand})$** , it highlights the most strategically valuable skills for both job seekers and hiring managers.

What I Learned ?

Through this project, I gained practical experience in the following areas:

- **SQL Mastery:** Improved my ability to write advanced SQL queries involving CTEs, joins, aggregations, filters, and mathematical functions for data analysis.
- **Data Cleaning & Transformation:** Learned to handle missing salary data and standardize hourly and annual salaries for fair comparison across roles.
- **Analytical Thinking:** Developed a strong understanding of how to extract actionable insights from raw datasets by asking the right questions and applying relevant metrics.
- **Insight Prioritization:** Discovered that combining salary data with skill demand (using a logarithmic weighting function) provides a more balanced and practical view of job market value than looking at either metric alone.
- **Data Storytelling:** Improved my ability to present findings through clear visualizations and structured explanations to communicate technical insights to a broader audience.

Conclusion

This analysis provides a data-driven roadmap for aspiring and current Data Analysts to strategically upskill based on market realities. Here are the key takeaways:

- SQL, Python, and Excel remain fundamental and frequently demanded in top-paying roles.
- Power BI and Tableau are common in roles with both high salary and high demand, making them practical tools to master.
- Skills like AWS and Snowflake, though lower in job frequency, command higher average salaries—suggesting niche but high-value specialization paths.
- By applying a combined metric of salary and demand, we identified the most strategically optimal skills to focus on — balancing high earning potential with broad job availability.

In an ever-evolving job market, these insights can guide personal development, hiring strategies, and educational content creation to better align with current industry needs.