Data Analysis Portfolio

Meriem
Paris, 05 August 2025
LinkedIn: https://www.linkedin.com/in/meriem-coeurdevey-87a0a835b/

Insights & Patterns: Synthetic projects demonstrating key data analysis skills.

Contents
1. Inside Airbnb: Pricing & Trends in Paris
2. How TikTok Works: What Drives Reach in 2025
3. Why Customers Leave: A Data Driven Churn Analysis
4. What Recruiters Want: A Data Analyst Job Market Study

---

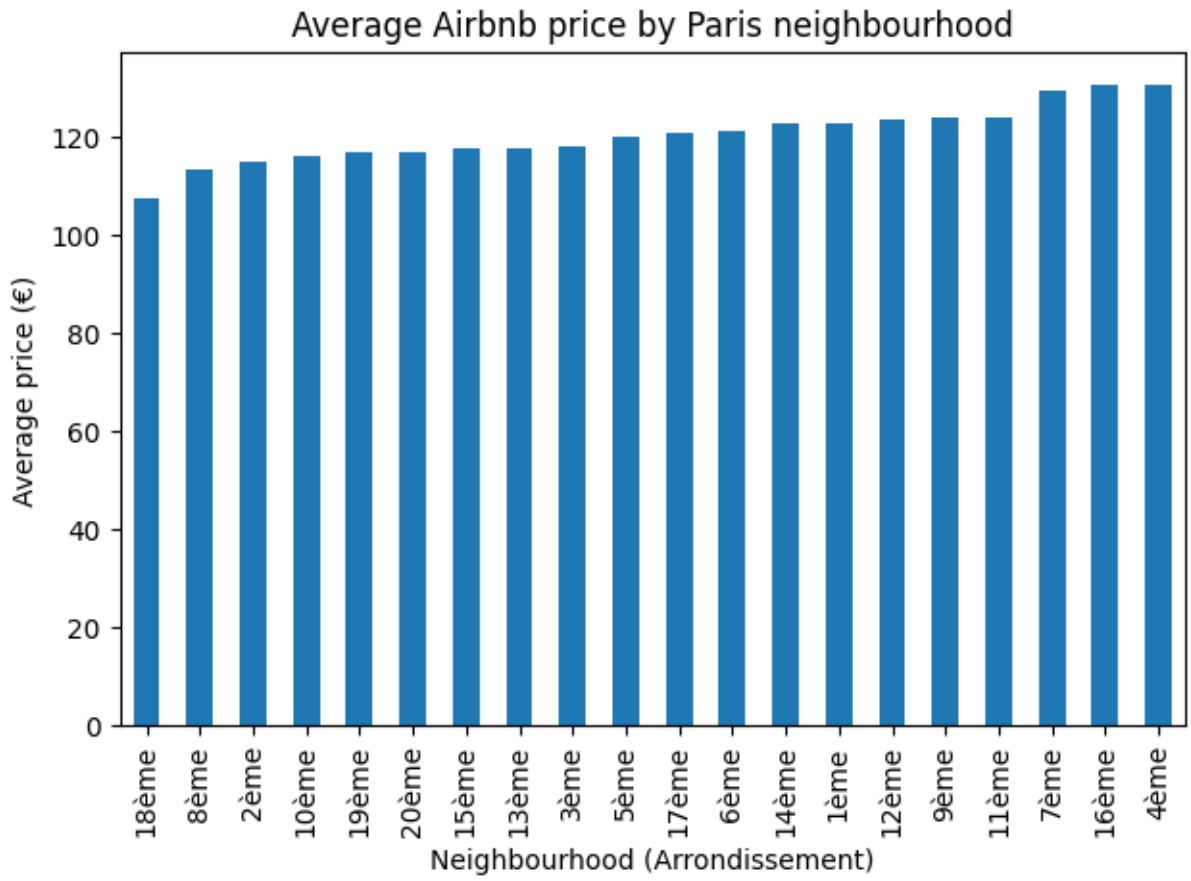
1. Inside Airbnb: Pricing & Trends in Paris

Overview:
This project explores a synthetic Airbnb dataset of 500 listings across the 20 arrondissements of Paris. Each listing includes the neighbourhood code, room type, nightly price, minimum nights and number of reviews.

Methodology & Analysis:
We generated the dataset randomly and then grouped the listings by neighbourhood to calculate the average price in euros. A bar chart visualises the variation of prices across districts, revealing differences between more expensive and more affordable areas.

Key Findings:
• Average prices vary significantly across arrondissements.
• Entire apartments tend to be priced higher than private or shared rooms.
• Even with synthetic data, grouping and aggregation techniques reveal meaningful patterns.

Skills Demonstrated:
Data generation, grouping and aggregation



, bar chart visualisation.

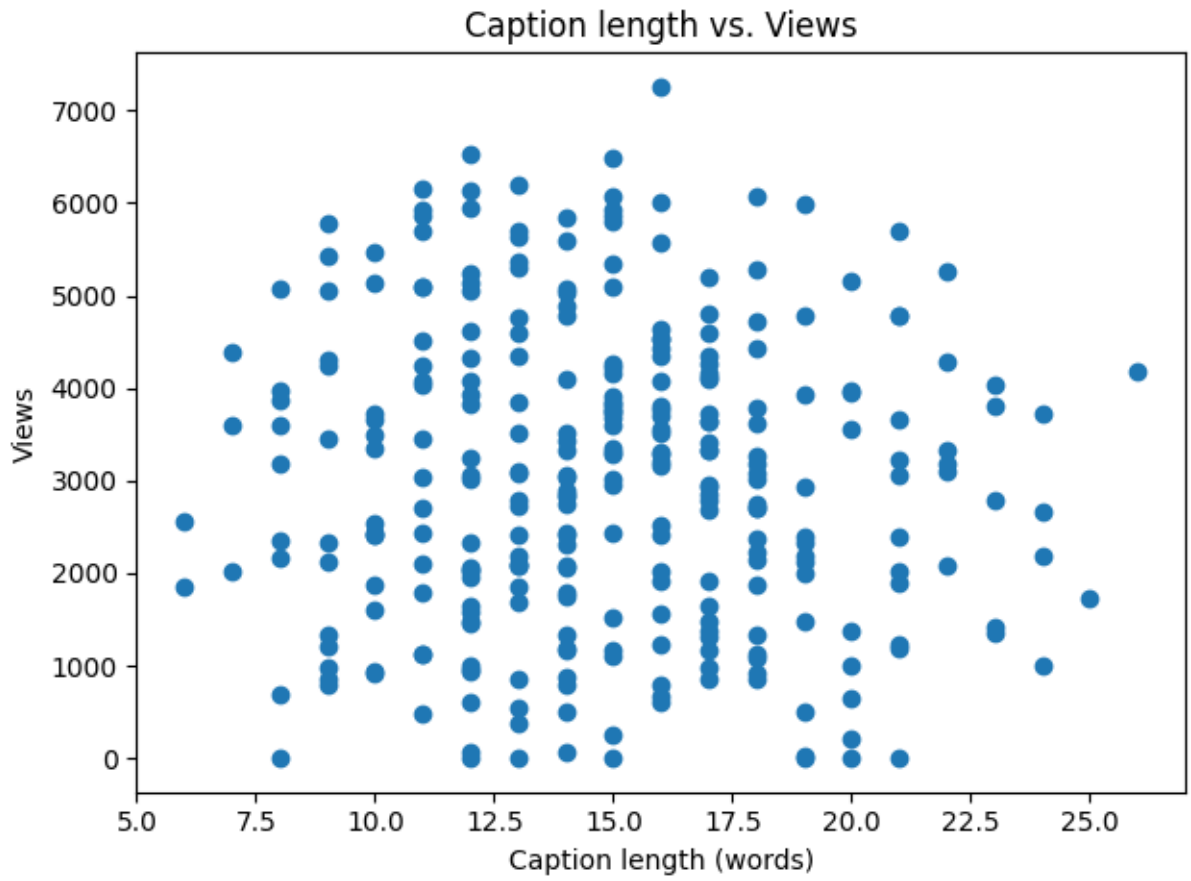
---

2. How TikTok Works: What Drives Reach in 2025

Overview:
This analysis uses a synthetic dataset of 300 TikTok videos to examine factors that may influence view counts. Features include video length, caption length, time of day posted, the use of trending sounds, number of hashtags and follower count.

Methodology & Analysis:
We synthesised view counts using a simple formula that rewards trending sounds and hashtags and penalises overly long videos. We then calculated the correlation between caption length and views and produced a scatter plot to visualise the relationship.

Key Findings:
• Caption length has a weak positive correlation with views.
• Trending sounds and multiple hashtags provide strong boosts to reach.
• Understanding how algorithmic factors interact helps optimise content strategy.



Skills Demonstrated:
Data simulation, correlation analysis, scatter plotting.

---

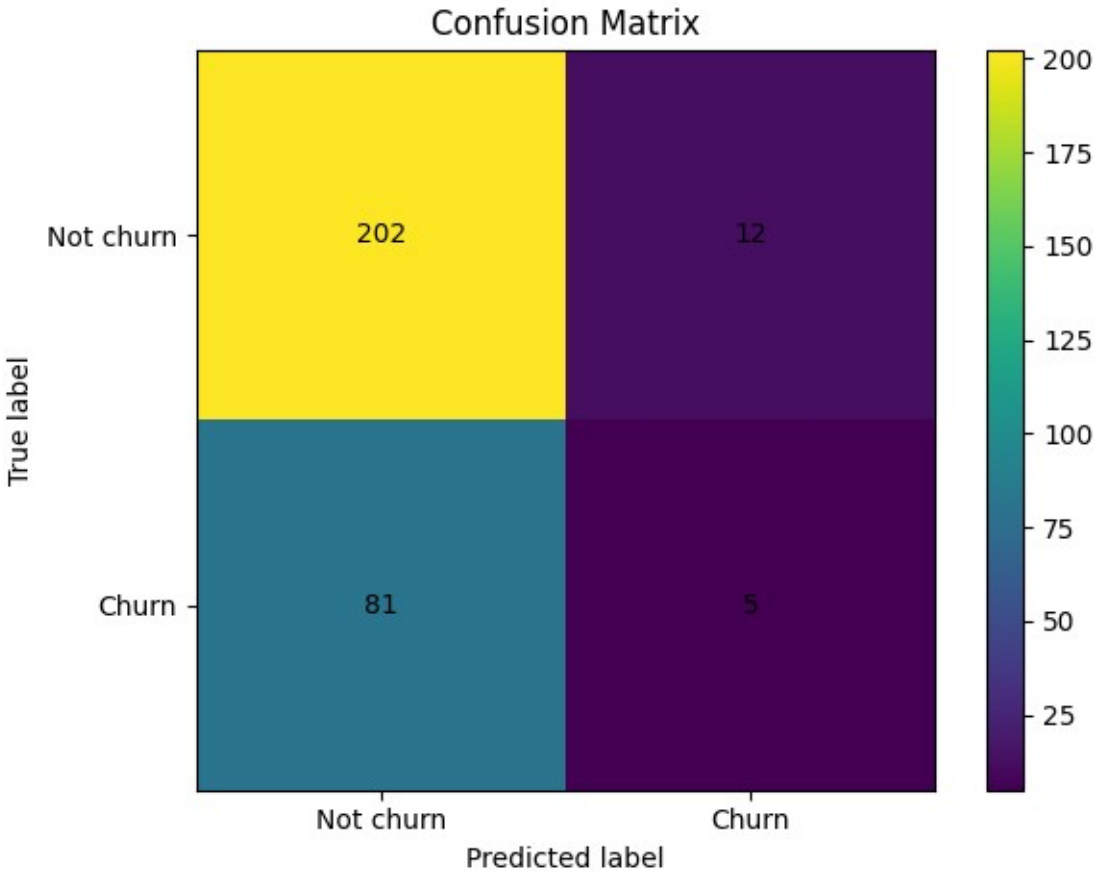
3. Why Customers Leave: A Data Driven Churn Analysis

Overview:
This project generates a synthetic telecom customer dataset with 1 000 records. Each customer has attributes such as tenure in months, monthly and total charges, contract type, payment method and internet service. A churn indicator marks whether they have left the service.

Methodology & Analysis:
We preprocess the data with one hot encoding for categorical variables and split it into training and test sets. A logistic regression model predicts churn, and a confusion matrix evaluates its performance.

Key Findings:
• Customers on month-to-month contracts and those with shorter tenure are more likely to churn.
• The logistic regression baseline achieves reasonable accuracy on the synthetic data.
• The confusion matrix helps interpret prediction successes and failures.

Skills Demonstrated:
Feature encoding, logistic regression modelling, evaluation metrics, confusion matrix plotng.



---

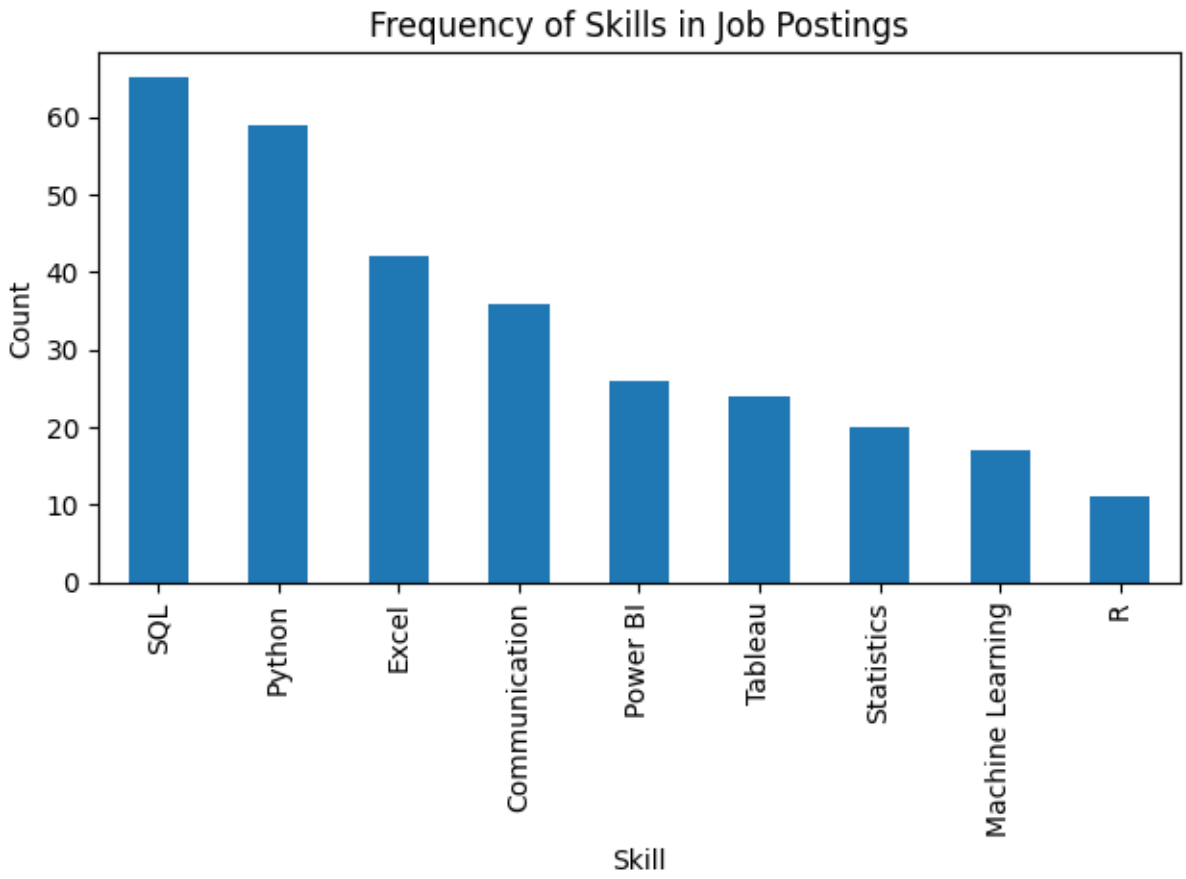
4. What Recruiters Want: A Data Analyst Job Market Study

Overview:
Our final project simulates 300 job postings for data-related roles. Each posting lists a job title, one required skill, an experience level and a salary estimate.

Methodology & Analysis:
We count how many times each skill appears in the postings and plot the frequencies to identify which tools and abilities are most in demand.

Key Findings:
• Python and SQL are the most frequently requested skills.
• Tools such as Excel, Power BI and Tableau remain important for data visualisation.
• Soft skills like communication and statistics also feature prominently.

Skills Demonstrated:
Text analysis, va



lue counting, bar chart plotting.

Conclusion
This portfolio brings together four mini-projects that span exploratory data analysis, synthetic data generation, predictive modelling and market research. Even with synthetic data, it demonstrates the core skills recruiters look for in entry-level data analysts: an ability to wrangle data, uncover patterns, build models and communicate insights visually.

Feel free to explore the accompanying Jupyter notebooks for full code and explanations.