

DATA ANALYST VS DATA SCIENTIST

The real difference, with a Swiggy example, then the exact steps and free resources for each path.

LOOKS AT THE PAST

ANALYST

"How many orders did we get last Monday?"

- Finds patterns in old data
- Builds reports and dashboards
- Tools: Excel, SQL, Power BI / Tableau, some Python

LOOKS AT THE FUTURE

SCIENTIST

"How many orders will we get next Monday?"

- Teaches a machine to predict
- Builds and tests ML models
- Tools: Python, scikit-learn, TensorFlow, stats

Same data, different question. The analyst explains what happened. The scientist predicts what will happen.

HOW A DATA SCIENTIST WORKS

1

PROBLEM DEFINE

Start with a business question, like "why are users leaving our app?" (customer churn).

2

EDA (EXPLORE THE DATA)

Find patterns, stats and anomalies in the data.

`pandas`

`Jupyter`

`matplotlib`

3

FEATURE ENGINEERING

Create the right inputs for the model. Better inputs = better model.

`NumPy`

`pandas`

4

BUILD THE ML MODEL

Train a model that learns from past data.

[scikit-learn](#)

[TensorFlow](#)

5

TEST, THEN DEPLOY

Check accuracy on data the model has never seen. If it's good enough, deploy it.

6

BUSINESS INSIGHTS

Turn predictions into action, e.g. find the customers with the highest churn risk and retain them first.

LEARNING PATHS (FREE)

ANALYST

- Excel basics
- SQL — [sqlbolt.com](#), [kaggle.com/learn](#) (Intro to SQL)
- Python + pandas — [kaggle.com/learn](#)
- A BI tool — Power BI, Tableau Public or Apache Superset
- Data storytelling: explain the chart in one line

SCIENTIST

- Stats and probability basics
- Python, NumPy, pandas
- ML — [scikit-learn.org](#), [kaggle.com/learn](#)
- Deep learning — [tensorflow.org](#), [pytorch.org](#)
- Deploy a model (a simple web app or API)

PORTFOLIO PROJECTS

Analyst: sales dashboard from a public dataset · food-delivery orders analysis · a SQL case study

Scientist: customer churn predictor · demand (orders) forecast · house price model

Publish each one on GitHub with a README and 2–3 screenshots. Kaggle has free datasets for all of these.

Which one first? If you like finding the story in data, start with analyst. If you like maths and building models, aim for scientist. Analyst skills (SQL, pandas) are useful on both paths.