* Wieviele Flüge haben die airlines hinter sich (absteigend)
* Wieviele Flüge kommen aus den einzelnen Städten (beliebteste) und in welche Stadt kommen sie an (beliebteste)
* Welche airline ist in welcher stadt öfter vertreten?
* Welche Flugroute ist die beliebteste? (von welcher stadt zu welcher stadt?)
* Wann starten die Flüge am meisten und wann kommen diese an? (airline-spezifisch)
* Wieviele stops werden gemacht (ist es spezifisch zur Flugroute? Welche airline macht mehr stops?)
* Welche klasse wird am meisten gewählt (airline spezifisch auch machen)
* Duration (wie lange dauern die Flüge von den Flugrouten (welche dauern am längsten? Welche ist die schnellste? Hängt es von airline ab? Hängt es von der Klasse ab? Wie lange dauert es mit stops?)
* Wie hoch sind die Preise durchschnittlich (bei welcher airline sind die Tickets billiger/teurer? Wie unterscheidet es sich bei den klassen? Welche flugrouten sind teurer/günsstiger, wie hängt es mit days\_left zusammen? Ist es von den städten abhängig? Von der startzeit?

Zusätzliche Fragen:

* in welcher Stadt starten die meisten Flüge?: 1.Delhi 2.Mumbai 3.Bangalore; in welcher Stadt landen die meisten Flieger: 1.Mumbai 2.Delhi 3.Bangalore -> Delhi, Mumbai und Bangalore sind bei beiden in den Top 3, aber aus welchem Grund? Warum starten die meisten Flüge in diesen Städten und warum landen die meisten Flieger wieder in diesen Städten?
* Warum starten die meisten flüge morgens, frühen morgens oder am abend (hat es etwas mit dem Preis zu tun?
* Warum ist Vistara die beliebteste airline in Indien?
* Entfernung der Städte voneinander (wegen duration)
* Welche airline, sollte ich nehmen wenn ich so schnell wie möglich an meinem ziel sein möchte?
* Warum ist der durchschnittliche Economy-Preis (wenn man alle airlines miteinbezieht) höher als der Business-preis??

Super—hier ist ein klarer **Schritt-für-Schritt-Plan** von jetzt bis zum GitHub-Portfolio. Einfach von oben nach unten abarbeiten.

**1) Power BI finalisieren**

1. **„h“ anhängen (Duration)**
   * Feld/Measure markieren → **Format**: *Benutzerdefiniert* → **Benutzerdefiniertes Format**:  
     0.00" h"
   * Alternative als Measure (nur Anzeige):
   * Avg\_Duration\_Display = FORMAT(AVERAGE('airlines'[duration]), "0.00") & " h"
2. **Preis-Beschriftung vereinheitlichen**
   * Für die KPI/Charts (nur Anzeige): 0.00" Tsd."
3. **Alt-Text setzen**
   * Visual auswählen → **Format** → **Allgemein** → **Alt-Text**:
     + Karte: „Karte der wichtigsten Städte (Delhi, Mumbai, Bangalore, Hyderabad, Chennai, Kolkata) mit Präsenz der Fluggesellschaften.“
     + Balken „Flüge pro Airline“: „Vergleich der Fluganzahl je Airline, Vistara höchste Werte.“
     + Balken „Ø-Preis je Airline“: „Durchschnittlicher Ticketpreis nach Airline.“
     + KPI Duration: „Durchschnittliche Flugdauer in Stunden.“
4. **Screenshots exportieren**
   * Gesamt-Dashboard + ggf. Detailseite → als PNG in /images/ speichern (z. B. dashboard\_screenshot.png).

**2) SQL sauber organisieren**

1. **Queries thematisch trennen** und kommentieren:
   * sql/01\_cleaning.sql
   * sql/02\_eda.sql
   * sql/03\_routes.sql
2. **Wichtige Beispiel-Query aufnehmen** (Routen ohne Doppelzählung):
3. SELECT
4. LEAST(source\_city, destination\_city) AS city\_a,
5. GREATEST(source\_city, destination\_city) AS city\_b,
6. ROUND(AVG(duration), 2) AS avg\_duration\_in\_h
7. FROM airlines
8. GROUP BY city\_a, city\_b
9. ORDER BY avg\_duration\_in\_h;
10. **Kurze README-Sätze** zu jeder Datei (was beantwortet wird).

**3) Datenordner vorbereiten**

* data/raw\_data.csv (falls erlaubt)
* data/cleaned\_data.csv (optional)
* Wenn zu groß/urheberrechtlich eingeschränkt: nur **Schema & Datendictionary** + Hinweis, woher die Daten stammen.

**4) README erstellen/ergänzen**

1. Struktur wie besprochen (Überblick, Workflow, Insights, Visualisierung, Dateien, **„Hinweis zu Preisen“**).
2. Den **Währungshinweis** einfügen:

Die Preisangaben enthalten keine Währungseinheit … Werte werden als reine Zahlen („Tsd.“) dargestellt.

1. Screenshot einbinden:
2. ![Power BI Dashboard – Indische Flugdaten](images/dashboard\_screenshot.png)

**5) Repository anlegen & befüllen**

flight-data-analysis/

├─ data/

├─ sql/

├─ powerbi/

├─ images/

└─ README.md

* **PBIX** nach /powerbi/ legen. Ist die Datei >100 MB → Git LFS nutzen oder extern verlinken (im README vermerken).
* Sinnvolle Commits: „add SQL EDA queries“, „add dashboard screenshot“, „add price currency note“, …

**6) Qualität & Zugänglichkeit**

* **Alt-Texte** (gemacht), **einheitliche Zahlentrenner**, **Legenden/Titel** prüfen.
* In README kurz die **Top-Insights** aufzählen (3–5 Bullet Points).

**7) Zukunft vorbereiten**

* Im README Abschnitt **„Zukünftige Erweiterungen“** (Web-Scraping, ML-Preisprognose …).
* In GitHub **Issues** anlegen: „Web-Scraper bauen“, „Preis-Forecast-Modell“, „Saisonanalyse“.

Wenn du willst, kopiere ich dir jetzt eine **Mini-Vorlage** für eine einzelne SQL-Analyse (Query + 1-Satz-Insight), die du für jede Auswertung wiederverwenden kannst:

### Analyse: Beliebteste Airline

```sql

SELECT airline, COUNT(\*) AS flights

FROM airlines

GROUP BY airline

ORDER BY flights DESC

LIMIT 5;

**Insight:** Vistara hat die meisten Flüge, gefolgt von Air India und Indigo.

Das war’s – mit dieser Checkliste bist du schnell „GitHub-ready“.

Billigste ticket buchen

Komfortabelste ticket buchen(z.b. nachts, weil wenige Flugzeuge fliegen und es dadurch ruhiger ist, business class

Hauptsache schnell