

# Airline BI Database – Phase 4 Query Catalog  
## Business Question Mapping (Queries 1–15, with Sample Outputs)

Each section documents:

- **\*\*Purpose\*\*** – Business question the query answers
- **\*\*Inputs\*\*** – Tables, key columns, and parameters
- **\*\*Outputs\*\*** – Result grain + description **\*\*with sample values from the actual query output\*\*** in ``03_analytics_queries.ipynb``
- **\*\*BI Value\*\*** – How the query supports analytics and decision-making

---

## 1) Top 10 busiest airports (arrivals + departures)

**\*\*Purpose\*\***

Identify the airports with the highest combined arrival and departure volume across all flights.

**\*\*Inputs\*\***

- ``airline.flights``
  - ``origin_airport_id`, `destination_airport_id`, `flight_date``
- ``airline.airports``
  - ``airport_id`, `iata_code`, `name`, `city`, `country``

**\*\*Outputs\*\***

Grain: **\*\*airport\*\***

Columns (sample from output):

airport_iata	airport_name	total_departures	total_arrivals	
total_movements				
-----	-----	-----	-----	-----
YCK	Colville Lake Airport	6	3	9
IBP	Iberia Airport	5	3	8
AZA	Phoenix-Mesa-Gateway Airport	3	5	8
GLV	Golovin Airport	1	6	7
PNA	Pamplona Airport	4	3	7

**\*\*BI Value\*\***

Highlights the main operational hubs in the network. These airports are candidates for additional gate capacity, lounge space, staffing, and also represent focal points for delay propagation.

---

## 2) Airline on-time performance summary (using BTS ``flight_performance``)

**\*\*Purpose\*\***

Summarize operational performance per airline: on-time percentage, delays, and cancellations/diversions.

**\*\*Inputs\*\***

- ``airline.flight_performance``
  - ``airline_iata`, `airport_iata`, `arr_delay`, `dep_delay`, `cancelled`, `diverted`, `year`, `month``
- ``airline.airlines``
  - ``airline_id`, `iata_code`, `name``

```
**Outputs**
Grain: **airline (BTS carriers)**

Columns (sample from output):

| airline_name      | iata_code | total_arrivals | delayed_arrivals | cancelled_arrivals |
| pct_delayed |
|-----|-----|-----|-----|-----|
|-----|
| Frontier Airlines | F9        | 208,624        | 58,481           | 4,835             |
| 0.2803           |
| Air Wisconsin    | ZW        | 52,393         | 11,859           | 764               |
| 0.2263           |
| American Airlines| AA        | 984,306        | 252,485          | 15,252            |
| 0.2565           |
| JetBlue Airways  | B6        | 240,282        | 60,121           | 3,735             |
| 0.2502           |
| Allegiant Air    | G4        | 117,210        | 24,897           | 2,018             |
| 0.2124           |
```

**\*\*BI Value\*\***  
Enables performance scorecards and SLA reviews across airlines. Operations and commercial teams can quickly see which carriers are more reliable and which require attention or agreements around delay handling.

---

## 3) Monthly passenger counts (via bookings)

```
**Purpose**
Track demand trends and seasonality by aggregating passenger bookings per calendar month.

**Inputs**
- `airline.bookings`
  - `booking_id`, `passenger_id`, `booking_date`
- `airline.passengers`
  - `passenger_id`
```

```
**Outputs**
Grain: **month**
```

Columns (sample from output):

month_start	total_bookings	unique_passengers
2025-02-01	1,688	1,436
2025-03-01	3,403	2,472
2025-04-01	3,236	2,415
2025-05-01	3,422	2,504
2025-06-01	3,268	2,445

**\*\*BI Value\*\***  
Shows monthly demand patterns (growth, peaks, troughs). Supports forecasting of capacity, staffing, and revenue, and provides context for promotion performance.

---

## 4) Loyalty tier transitions (current vs miles-based target)

```
**Purpose**
Compare each member's current tier to the tier they qualify for based on miles, highlighting potential upgrades or downgrades.
```

```
**Inputs**
```

- `airline.loyalty\_accounts`
  - `loyalty\_id`, `passenger\_id`, `tier`, `miles\_balance` / `ytd\_miles`
- `airline.miles\_transactions`
  - `loyalty\_id`, `miles\_delta`, `txn\_date`, `txn\_type`

#### **\*\*Outputs\*\***

Grain: **\*\*(current\_tier, target\_tier)\*\*** summary counts

Columns (sample from output):

current_tier	target_tier	member_count
Basic	Basic	353
Basic	Gold	181
Basic	Platinum	73
Basic	Silver	138
Silver	Basic	350

#### **\*\*BI Value\*\***

Identifies members whose current tier is “behind” their earned miles (good upgrade candidates) and potential downgrades. This is vital for loyalty program management and targeted retention campaigns.

---

### **## 5) Revenue per fare class (bookings + payments)**

#### **\*\*Purpose\*\***

Understand revenue mix across fare classes (e.g., Basic, Standard, Flexible, Business, First).

#### **\*\*Inputs\*\***

- `airline.bookings`
  - `booking\_id`, `fare\_class`
- `airline.payments`
  - `booking\_id`, `amount\_usd`, `status`, `paid\_at`

#### **\*\*Outputs\*\***

Grain: **\*\*(fare\_class)\*\***

Columns (sample from output):

fare_class	num_bookings	total_revenue	avg_revenue_per_booking
Basic	13,903	1,572,721.97	113.12
Standard	11,827	1,338,850.26	113.20
Flexible	8,211	936,208.77	114.02
Business	4,029	458,256.95	113.74
First	2,030	233,756.91	115.15

#### **\*\*BI Value\*\***

Shows the relative revenue contribution of each fare product. Supports fare strategy, upsell tactics, and product design (e.g., whether to invest in Premium/Business cabins).

---

### **## 6) Ranking airlines by average delay**

#### **\*\*Purpose\*\***

Use a window function to rank airlines by mean delay.

#### **\*\*Inputs\*\***

- `airline.flight\_performance` / `airline.flights`
  - `airline\_id`, `arr\_delay` / `delay\_minutes`
- `airline.airlines`

### **\*\*Outputs\*\***

Grain: **\*\*airline\*\***

Columns (sample from output):

airline_name	iata_code	avg_delay_minutes	delay_rank
Red Jet Mexico	4X	287.00	1
Cargo Plus Aviation	8L	257.00	2
Sriwijaya Air	SJ	253.50	3
Armenian International Airways	MV	251.00	4
Malaysia Airlines	MH	226.33	5

### **\*\*BI Value\*\***

Quickly ranks carriers by punctuality, identifying worst offenders. Useful for operational negotiations, scheduling changes, and customer communications.

---

## **## 7) Running monthly revenue totals**

### **\*\*Purpose\*\***

Build a time series of revenue with a running cumulative total using window `SUM()`.

### **\*\*Inputs\*\***

- `airline.payments`  
- `amount`, `status`, `paid\_at` (filtered to successful statuses)

### **\*\*Outputs\*\***

Grain: **\*\*month\*\***

Columns (sample from output):

month_start	revenue	running_cumulative_revenue
2025-02-01	185,699.32	185,699.32
2025-03-01	383,880.42	569,579.74
2025-04-01	369,920.05	939,499.79
2025-05-01	389,381.51	1,328,881.30
2025-06-01	372,051.23	1,700,932.53

### **\*\*BI Value\*\***

Supports revenue pacing dashboards and comparison to budget/forecast over time. Clearly shows growth trajectory and the effect of seasonal peaks.

---

## **## 8) Percent of flights delayed by month**

### **\*\*Purpose\*\***

Measure the share of flights that are delayed each month.

### **\*\*Inputs\*\***

- `airline.flights` / `airline.flight\_performance`  
- `flight\_date`, `delay\_minutes` (or arrival/departure delay fields)

### **\*\*Outputs\*\***

Grain: **\*\*month\*\***

Columns (sample from output):

month_start	total_flights	delayed_flights	pct_delayed
2024-01-01	140	105	0.7500

2024-02-01	117	87	0.7436	
2024-03-01	144	119	0.8264	
2024-04-01	154	114	0.7403	
2024-05-01	125	99	0.7920	

#### **\*\*BI Value\*\***

Shows monthly reliability performance and reveals seasonality (e.g., winter weather). Useful for root-cause analysis and tracking the impact of process changes.

---

### **## 9) Customer lifetime value (CLV) window function**

#### **\*\*Purpose\*\***

Compute cumulative revenue per passenger over time using a CLV-style window function.

#### **\*\*Inputs\*\***

- `airline.bookings`  
 - `booking\_id`, `passenger\_id`  
 - `airline.payments`  
 - `booking\_id`, `amount`, `paid\_at`, `status`

#### **\*\*Outputs\*\***

Grain: **\*\*payment event per passenger\*\***, with cumulative CLV

Columns (sample from output for `passenger\_id = 1`):

passenger_id	paid_date	amount_usd	clv_to_date	
1	2025-03-10	90.98	90.98	
1	2025-04-09	73.00	163.98	
1	2025-05-04	121.78	285.76	
1	2025-07-25	74.34	360.10	
1	2025-08-29	168.50	528.60	

#### **\*\*BI Value\*\***

Provides a customer-level view of revenue over time, supporting segmentation into high-value vs. low-value customers and informing retention and marketing priorities.

---

### **## 10) Dense\_rank route distance analysis (distance computed on the fly)**

#### **\*\*Purpose\*\***

Rank the longest routes using approximate distances derived from airport coordinates and a window `DENSE\_RANK()`.

#### **\*\*Inputs\*\***

- `airline.routes`  
 - `route\_id`, `origin\_airport\_id`, `destination\_airport\_id`  
 - `airline.airports`  
 - `airport\_id`, `iata\_code`, `latitude`, `longitude`

#### **\*\*Outputs\*\***

Grain: **\*\*route\*\***

Columns (sample from output):

route_id	origin_iata	destination_iata	distance_nm	distance_rank	
2781	NLK	TLA	20,839.17	1	
2583	HOM	KTF	20,367.45	2	
3884	UVE	MCG	19,970.83	3	
4006	KTS	BHS	19,870.81	4	
3220	KSM	FRE	19,824.93	5	

### **\*\*BI Value\*\***

Highlights the longest segments in the network, which often drive distinct cost and product considerations (fuel, crew duty time, cabin product). Useful for fleet assignment and long-haul strategy.

----

### **## 11) Airport connectivity graph from busiest origin**

#### **\*\*Purpose\*\***

Use a recursive CTE to find all airports reachable from the busiest origin within up to 3 hops.

#### **\*\*Inputs\*\***

- `airline.routes`  
- `origin\_airport\_id`, `destination\_airport\_id`  
- `airline.airports`  
- `airport\_id`, `iata\_code`

#### **\*\*Outputs\*\***

Grain: **\*\*origin-destination-hop combination\*\***

Columns (sample from output):

origin_iata	dest_iata	hops	path
YCK	EIK	1	[YCK, EIK]
YCK	NVT	1	[YCK, NVT]
YCK	NYR	1	[YCK, NYR]
YCK	PIP	1	[YCK, PIP]
YCK	RUM	1	[YCK, RUM]

### **\*\*BI Value\*\***

Shows the reach of a key hub and the set of airports that can be served directly or via one connection. Supports hub planning, connection design, and network optimization.

----

### **## 12) Multi-hop routes: detailed paths up to 3 hops from busiest origin**

#### **\*\*Purpose\*\***

List explicit multi-hop routes (up to 3 hops) from the busiest origin, showing full paths.

#### **\*\*Inputs\*\***

- `airline.routes`  
- `airline.airports`

#### **\*\*Outputs\*\***

Grain: **\*\*multi-hop path\*\*** from origin to destination

Columns (sample from output):

origin_iata	dest_iata	hops	path
YCK	AHS	3	[YCK, NVT, YCW, AHS]
YCK	AKI	3	[YCK, NVT, YCW, AKI]
YCK	BTT	3	[YCK, RUM, FEN, BTT]
YCK	HEL	3	[YCK, RUM, TPP, HEL]
YCK	YJF	3	[YCK, TJB, FUK, YJF]

### **\*\*BI Value\*\***

Provides concrete connection options and reveals how complex some journeys are (e.g., 3-leg itineraries). Supports decisions on adding direct routes or retiming flights to improve connection quality.

---

## 13) Payment success rate by booking channel (Captured + Authorized as success)

**\*\*Purpose\*\***

Evaluate payment performance by booking channel, treating `Captured` and `Authorized` as successful outcomes.

**\*\*Inputs\*\***

- `airline.bookings`  
- `booking\_id`, `booking\_channel`  
- `airline.payments`  
- `booking\_id`, `status`

**\*\*Outputs\*\***

Grain: **\*\*booking\_channel\*\***

Columns (sample from output):

booking_channel	total_payments	successful_payments	success_rate
Mobile	10,088	8,101	0.8030
Web	21,919	17,514	0.7990
Call Center	3,942	3,126	0.7930
Travel Agent	4,051	3,212	0.7929

**\*\*BI Value\*\***

Highlights differences in conversion between channels. A lower success rate on a specific channel (e.g., Web) can indicate technical issues or UX friction that directly reduce revenue.

---

## 14) Worst routes by delay + cancellations (no volume cutoff)

**\*\*Purpose\*\***

Identify the most problematic routes by combining average delay and cancellation rate, with no minimum volume filter.

**\*\*Inputs\*\***

- `airline.flights`  
- `route\_id`, `delay\_minutes`, `status`  
- `airline.routes`  
- `airline.airports`

**\*\*Outputs\*\***

Grain: **\*\*route\*\***

Columns (sample from output):

route_id	origin_iata	destination_iata	total_flights	avg_delay_minutes	cancel_rate
3107	LHA	RIA	1	300.0	1.0
845	OCV	ZVK	1	300.0	1.0
2065	MYP	PAS	1	300.0	1.0
4085	CRQ	SAA	1	300.0	1.0
1449	BPY	GJT	1	299.0	1.0

## **\*\*BI Value\*\***

Provides a route-level “watch list” for operational remediation. Even with synthetic data, this pattern supports a dashboard tile that flags routes with extreme delay and cancellation metrics.

----

## **## 15) High-value loyalty members (top 5% by lifetime miles)**

### **\*\*Purpose\*\***

Use window functions (e.g., `CUME_DIST()` / `PERCENT_RANK()`) to identify the top 5% of members by lifetime miles.

### **\*\*Inputs\*\***

- ``airline.loyalty_accounts``
  - ``loyalty_id``, ``passenger_id``, ``tier``, ``miles_balance``
- ``airline.miles_transactions``
  - ``loyalty_id``, ``miles_delta``, ``txn_date``

### **\*\*Outputs\*\***

Grain: **\*\*loyalty account\*\***

Columns (sample from output):

<code>loyalty_id</code>	<code>passenger_id</code>	<code>tier</code>	<code>miles_balance</code>	<code>lifetime_miles</code>	<code>percentile_rank</code>
1385	2298	Gold	40,763	218,556	1.0000
1536	2543	Basic	41,192	215,170	0.9997
649	1065	Silver	6,116	210,018	0.9993
1714	2842	Gold	58,618	202,778	0.9990
642	1047	Basic	22,748	197,384	0.9987

## **\*\*BI Value\*\***

Enables a focused VIP strategy: these members can be targeted for special offers, dedicated support, and retention programs, maximizing the value of the loyalty program.

----

\_End of Phase 4 Business Question Mapping with actual query outputs from  
`03\_analytics\_queries.ipynb`.\_