# Expansion: The 2 Goals and Scope of this Project

**Association Mining for Targeted Marketing:** Identify traveler behavior patterns by uncovering correlations between airports and preferred airlines, aiding airline marketing teams in tailoring strategies for route planning and passenger engagement.

**Graph-Based Analysis for Delay Reduction:** Utilize delay-integrated network representation to analyze delay propagation across airports and airlines, helping airline operations teams identify influential nodes and routes to minimize delays and enhance operational efficiency.

# Association Mining for Targeted Marketing (2022 Data): Using FP-Growth

- Association rule mining, like FP-Growth used here, emphasizes discovering frequent patterns within transactions, beneficial for **identifying strategic airline route planning** rather than directly pinpointing airlines with the lowest delays for specific routes.
- Specifically, we filter and analyze frequent itemsets and association rules based on **high lift and support,** centered on departures from ORD or MDW, aiding airlines in optimizing flight schedules and operations.

| antecedent | consequent | confidence | lift | support |
|---|---|---|---|---|
| [GoJet Airlines, ... | [ LLC d/b/a Unite... | 1.0 | 69.78315290933695 | 0.01433010631232457 |
| [ORD] | [American Airline... | 0.16698256824978489 | 1.2412164316960606 | 0.13453138710193477 |
| [ORD] | [Republic Airlines] | 0.05270441841014255 | 1.2412164316960606 | 0.04246190838621479 |
| [ORD] | [SkyWest Airlines... | 0.18223610183463604 | 1.2137246819820124 | 0.14682056825949322 |
| [ORD] | [Air Wisconsin Ai... | 0.09171976821849559 | 1.2412164316960606 | 0.07389506445154378 |
| [ORD] | [United Air Lines... | 0.22045718480543472 | 1.2412164316960606 | 0.17761381430004994 |
| [ORD] | [Envoy Air] | 0.13953222497006457 | 1.2412164316960606 | 0.11241570882154751 |
| [LGA, ORD] | [American Airline... | 0.41333752752437875 | 3.0724244834492924 | 0.012740026856829827 |
| [ LLC d/b/a Unite... | [GoJet Airlines] | 1.0 | 69.78315290933695 | 0.01433010631232457 |
| [American Airline... | [LGA] | 0.0946992901156715 | 2.6182113477778586 | 0.012740026856829827 |
| [American Airline... | [DFW] | 0.07567294872256855 | 4.7489322146464 | 0.01018038675774073 |
| [DFW, ORD] | [American Airline... | 0.6578947368421053 | 4.890269482939448 | 0.01018038675774073 |
| [MDW] | [Southwest Airlin... | 0.9315006984633806 | 4.520515517299385 | 0.18102666776550205 |

# Association Mining for Targeted Marketing (2022 Data): Outcomes for ORD, MDW, and Connections

- **MDW (Chicago Midway International Airport) and Southwest Airlines:** The association between MDW and Southwest Airlines reveals a very high confidence level (93.15%) and a significant lift value (4.52), indicating an extremely strong connection between Southwest Airlines and flights departing from MDW.

- **Strong Association with ORD (Chicago O'Hare Airport): American Airlines** demonstrates a moderate confidence level (16%) with ORD, indicating a significant association with the airport, along with other airlines like SkyWest Airlines and United Airlines. The high lift values (>1) suggest these airlines are more frequently associated with ORD than expected randomly.

- **Connection between Airports (e.g., ORD and LGA):** The association rule between LGA (LaGuardia Airport) and ORD (Chicago O'Hare Airport) with **American Airlines** shows a high confidence level (41.3%), indicating a strong association between these airports in terms of flights operated by American Airlines. The notably high lift value (3.07) indicates a substantial influence or relationship between American Airlines' flights from LGA to ORD.

# Association Mining for Targeted Marketing (2022 Data): Potential Opportunities for Expansion

**Leveraging American Airlines' Strong Presence at ORD to Attract Family Travelers to Chicago:**

With 600+ daily departures and access to 150+ global destinations (Source: dataset), American Airlines leads at ORD, making it an optimal choice for families planning trips to Chicago.

A 2021 Expedia Group report highlighted that 70% of families seek comprehensive travel packages including accommodations, activities, and transport—showcasing a growing trend in family travel preferences.Chicago is a popular tourist destination, with over 55 million visitors in 2022. This high level of tourism highlights the city's appeal as a family-friendly destination.

American Airlines can partner with key Chicago attractions like Millennium Park, Navy Pier, and Lincoln Park Zoo—highly favored spots for families visiting the city. Collaborating with GOCHICAGO, renowned for all-inclusive passes encompassing attractions, transportation, and dining discounts, American Airlines can offer exclusive deals, enhancing the appeal for families planning Chicago visits.



# Why Graph Theory may be better used for customer segments like consultants

- For customers segments that want to focus on **less delays** and **more route options** (like consultants), instead of association mining, we can incorporate **graph theory**.

## Enhanced Approach Through Graph Theory:

- Constructing a graph with nodes representing airports and edges indicating connections, leveraging shared passengers or historical delays.
- Incorporating Delay Information: **Assigning weights to edges based on historical delay data (Departure + Arrival)**

| src | dst | weight |
|---|---|---|
| SPI | ORD | 28.591269841269842 |
| SPI | ORD | 28.591269841269842 |
| SPI | ORD | 28.591269841269842 |
| SPI | ORD | 28.591269841269842 |
| ATL | GSP | 19.09255079006772 |
| ATL | GSP | 19.09255079006772 |
| DAL | LGB | 28.321782178217823 |
| MDW | MEM | 52.754666666666665 |
| MDW | MEM | 52.754666666666665 |
| PHL | MCO | 39.9860594795539 |
| PHL | MCO | 39.9860594795539 |
| SJC | LIH | 16.50570342205323 |

| Airline | average_delay |
|---|---|
| Envoy Air | 19.823666393548663 |
| Air Wisconsin Air... | 22.084891425572394 |
| GoJet Airlines, L... | 22.74120433017591 |
| Republic Airlines | 26.74928644822468 |
| Endeavor Air Inc. | 27.708588957055213 |
| Delta Air Lines Inc. | 28.90657769304099 |
| United Air Lines ... | 29.224739341667124 |
| Alaska Airlines Inc. | 29.586942875078467 |
| SkyWest Airlines ... | 31.02334366524603 |
| Horizon Air | 34.01578947368421 |
| American Airlines... | 39.14738207632158 |

# 3 applications from graph theory that would be beneficial for American Airlines

**Exploration of Indirect Flight Paths (JFK to SFO) by Top Airlines**

- Consultants would benefit from this analysis if there are no direct flights between JFK and SFO as it helps identify popular indirect routes
- Airlines, by leveraging this analysis, can identify market gaps, optimize their flight schedules, and potentially introduce new routes based on the demand for indirect flights between these two airports.

```
# Find paths from JFK to SFO that aren't direct flights and get corresponding airlines
indirect_paths = graph_with_airline_info.find("(a)-[ab]->(b); (b)-[bc]->(c)") \
    .where(f"a.id = '{source_airport}'") \
    .where(f"c.id = '{destination_airport}'") \
    .where("ab.airline != bc.airline") \
    .selectExpr("a.id AS src", "b.id AS via", "c.id AS dst", "ab.airline AS airline1", "bc.airline AS airline2")
```

```
+---+---+---+--------------------+--------------------+----------+
|src|via|dst|            airline1|            airline2|path_count|
+---+---+---+--------------------+--------------------+----------+
|JFK|LAX|SFO|     JetBlue Airways|SkyWest Airlines ...|   5748750|
|JFK|LAX|SFO|American Airlines...|SkyWest Airlines ...|   5681298|
|JFK|LAX|SFO|Delta Air Lines Inc.|SkyWest Airlines ...|   4923996|
|JFK|LAX|SFO|     JetBlue Airways|United Air Lines ...|   2728125|
|JFK|LAX|SFO|American Airlines...|United Air Lines ...|   2696115|
+---+---+---+--------------------+--------------------+----------+
```

**BFS (Breadth-First Search) algorithm considering delays in a graph (representing airline connections)**

- The algorithm computes levels (or distances) between nodes (airports) while factoring in **delays** as weights on edges (routes).
- Levels are determined by incrementally adding delays to the current level as the algorithm traverses the graph, helping airlines understand the impact of delays on the connectivity and distances between airports in their network.

```
Node: JFK, Level: 0
Node: ORD, Level: 1
Node: SRQ, Level: 1
Node: BTV, Level: 1
Node: JAC, Level: 1
Node: ORF, Level: 1
Node: MIA, Level: 1
Node: MEI, Level: 2
Node: ART, Level: 2
Node: BRW, Level: 3
Node: ADK, Level: 3
Node: OME, Level: 3
```

# 3 Insights derived from graph theory for American Airlines

**Airline Route Optimization: Leveraging High Connectivity Airports for Enhanced Efficiency**

- In-Degree and Out-Degree:
  - In-degree: flights arriving at an airport. Out-degree: flights departing from an airport.
- Highly Connected Airport:
  - Example: Burbank Airport (BUR) shows a total degree of 35,051, indicating high connectivity due to its 17,549 incoming and 17,502 outgoing flights.
- Use Case for Airlines:
  - Airlines can utilize this information to optimize their flight schedules by focusing on routes connecting to airports with high connectivity.

| id | in_degree | out_degree | total_degree |
|---|---|---|---|
| ATY | 2 | 2 | 4 |
| BGM | 208 | 209 | 417 |
| BUR | 17549 | 17502 | 35051 |
| DLG | 229 | 228 | 457 |
| DRT | 383 | 385 | 768 |
| EUG | 5235 | 5231 | 10466 |
| FOD | 336 | 334 | 670 |
| GEG | 11160 | 11153 | 22313 |
| GRB | 2645 | 2644 | 5289 |
| GRR | 9720 | 9688 | 19408 |
| GTF | 1288 | 1283 | 2571 |
| IDA | 2692 | 2683 | 5375 |
| INL | 330 | 330 | 660 |
| JLN | 368 | 363 | 731 |
| LWB | 345 | 346 | 691 |
| MSY | 26905 | 26874 | 53779 |
| PPG | 25 | 25 | 50 |
| PSE | 413 | 415 | 828 |
| PSG | 412 | 410 | 822 |
| PVU | 641 | 641 | 1282 |