Rules

December 3, 2025

0.0.1 Chunk 1: Imports & Setup

We import `apriori` instead of `fpgrowth`.

```
[1]: from google.colab import drive
     drive.mount('/content/drive')
```

```
Mounted at /content/drive
```

```
[2]: # ==========================================
     # TASK 3 FINAL: WEATHER + APRIORI
     # ==========================================
     import warnings
     import pandas as pd
     import numpy as np
     import matplotlib.pyplot as plt
     import seaborn as sns
     from mlxtend.frequent_patterns import apriori, association_rules
     warnings.filterwarnings('ignore')
     # Configuration
     FLIGHT_DATA = '/content/drive/MyDrive/Colab Notebooks/term/US_flights_2023.csv'
     TIER_DATA = '/content/drive/MyDrive/Colab Notebooks/term/
      ↪airport_performance_tiers_enriched.csv'
     WEATHER_DATA = '/content/drive/MyDrive/Colab Notebooks/term/
      ↪weather_meteo_by_airport.csv'

     print("Libraries loaded. Algorithm set to: APRIORI.")
```

```
Libraries loaded. Algorithm set to: APRIORI.
```

0.0.2 Chunk 2: Loading & Merging (Same as before)

We still need to build the Master Dataset exactly the same way.

```
[3]: print("--- 1. Loading & Merging Master Dataset ---")

     # 1. Load Main Flight Data
     cols_flight = ['FlightDate', 'Airline', 'Dep_Airport', 'Dep_Delay']
     df = pd.read_csv(FLIGHT_DATA, usecols=cols_flight, low_memory=False)
```

```
# 2. Load Performance Tiers
df_tiers = pd.read_csv(TIER_DATA)
df_tiers = df_tiers[['Dep_Airport', 'Performance_Tier']]

# 3. Load Weather Data
df_weather = pd.read_csv(WEATHER_DATA)
cols_weather = ['time', 'airport_id', 'prcp', 'snow', 'wspd', 'tmin']
df_weather = df_weather[cols_weather]

print("... Merging Datasets ...")

# Merge Tiers
df_merged = df.merge(df_tiers, on='Dep_Airport', how='left')

# Merge Weather (Match Date + Airport)
df_merged = df_merged.merge(
    df_weather,
    left_on=['FlightDate', 'Dep_Airport'],
    right_on=['time', 'airport_id'],
    how='left'
)

# Clean up
df_merged.drop(columns=['time', 'airport_id'], inplace=True)
df_merged.dropna(subset=['Dep_Delay', 'Performance_Tier', 'wspd'], inplace=True)

print(f"Success. Master Dataset created: {len(df_merged):,} flights.")
```

```
--- 1. Loading & Merging Master Dataset ---
… Merging Datasets …
Success. Master Dataset created: 6,678,080 flights.
```

0.0.3 Chunk 3: Feature Engineering (Risk Factors)

We create the boolean flags for Apriori to consume.

[4]:

```
print("\n--- 2. Creating Risk Factors ---")

# 1. Map Tier Names
tier_mapping = {
    0: 'Tier_0_Secondary',
    1: 'Tier_1_HighRisk',
    2: 'Tier_2_Underperforming',
    3: 'Tier_3_Efficient',
    4: 'Tier_4_MegaHub'
}
df_merged['Tier_Name'] = df_merged['Performance_Tier'].map(tier_mapping)
```

```
# 2. Discretize Delays (Target)
bins = [-np.inf, 14, 60, np.inf]
labels = ['OnTime', 'Late', 'Severe']
df_merged['Delay_Class'] = pd.cut(df_merged['Dep_Delay'], bins=bins,␣
 ↪labels=labels)

# 3. Create Weather Flags (True/False)
df_merged['Weather_Snow'] = df_merged['snow'] > 0
df_merged['Weather_Rain'] = df_merged['prcp'] > 5
df_merged['Weather_Wind'] = df_merged['wspd'] > 25
df_merged['Weather_Freezing'] = df_merged['tmin'] < 0

# 4. Select Final Columns
mining_df = df_merged[[
    'Airline',
    'Tier_Name',
    'Weather_Snow',
    'Weather_Rain',
    'Weather_Wind',
    'Weather_Freezing',
    'Delay_Class'
]].copy()

print("Feature Engineering Complete. Preview:")
print(mining_df.head())
```

```
--- 2. Creating Risk Factors ---
Feature Engineering Complete. Preview:
        Airline         Tier_Name  Weather_Snow  Weather_Rain  Weather_Wind  \
0  Endeavor Air  Tier_0_Secondary         False         False         False
1  Endeavor Air  Tier_0_Secondary         False          True         False
2  Endeavor Air  Tier_0_Secondary         False          True         False
3  Endeavor Air  Tier_0_Secondary         False         False         False
4  Endeavor Air  Tier_0_Secondary         False          True         False

   Weather_Freezing Delay_Class
0              True      OnTime
1              True      OnTime
2             False      OnTime
3             False      OnTime
4             False      OnTime
```

0.0.4 Chunk 4: Running Apriori (The Switch)

Warning: This step will take longer than `fpgrowth`.

[5]:

```python
print("\n--- 3. Running Apriori (Full Dataset) ---")

# 1. One-Hot Encoding (Required for Apriori)
# We strictly cast to 'bool' to minimize memory usage
df_ohe = pd.get_dummies(mining_df, prefix_sep='=').astype(bool)

# 2. Run Apriori Algorithm
# min_support=0.001: We look for patterns occurring in 0.1% of flights (~6k flights)
# low_memory=True: This is a pandas feature, but mlxtend relies on your RAM.
print("Mining Frequent Patterns (This may take 2-5 minutes)...")
frequent_itemsets = apriori(df_ohe, min_support=0.001, use_colnames=True)

# 3. Generate Rules
print("Generating Association Rules...")
rules = association_rules(frequent_itemsets, metric="lift", min_threshold=1.0)

print(f"Computation Complete. Found {len(rules)} association rules.")
```

```
--- 3. Running Apriori (Full Dataset) ---
Mining Frequent Patterns (This may take 2-5 minutes)…
Generating Association Rules…
Computation Complete. Found 2238 association rules.
```

0.0.5 Chunk 5: Analysis & Output

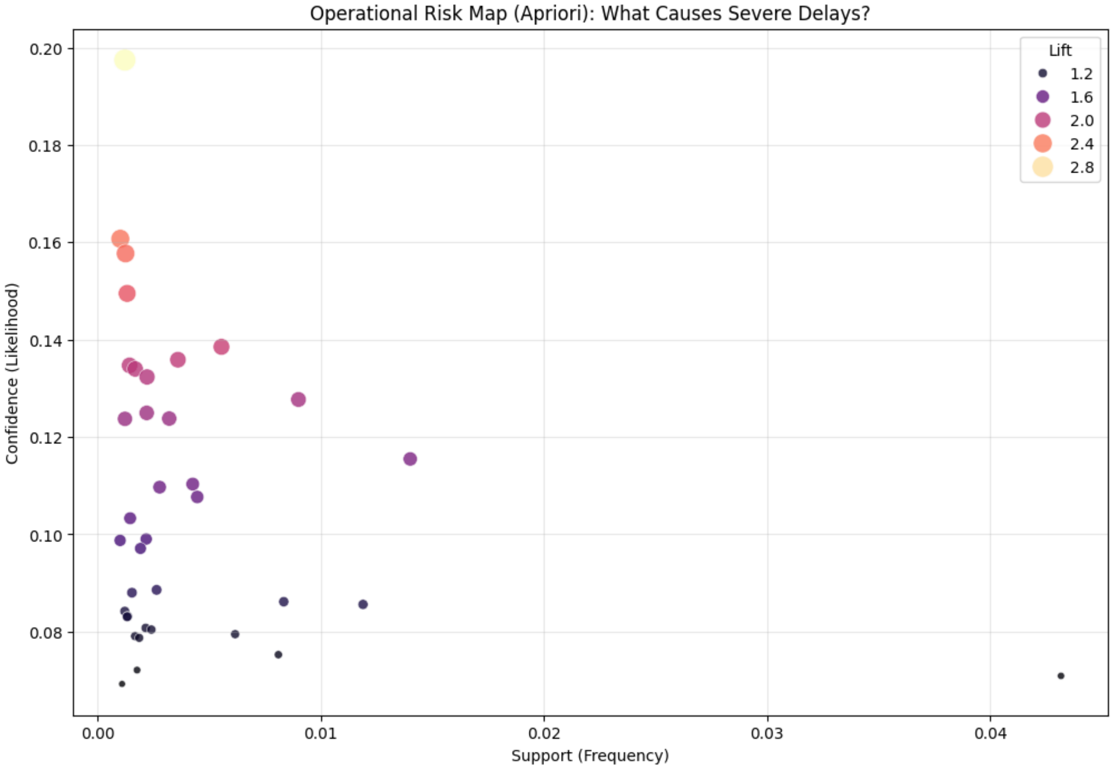
The analysis remains the same. We want to see who fails when it snows.

[6]:

```python
print("\n--- 4. Analysis: Predictors of Severe Delays ---")

# Filter: Consequent must be SEVERE Delay
target = {'Delay_Class=Severe'}
severe_rules = rules[rules['consequents'] == target].copy()

# Sort by Lift
severe_rules = severe_rules.sort_values('lift', ascending=False)

# Display Top 15
cols_show = ['antecedents', 'consequents', 'support', 'confidence', 'lift']
print("TOP RISK FACTORS (Apriori Results):")
print(severe_rules[cols_show].head(15))

# Visualization
plt.figure(figsize=(12, 8))
sns.scatterplot(
    x="support",
    y="confidence",
    size="lift",
    data=severe_rules,
    hue="lift",
    palette="magma",
    sizes=(20, 200),
    alpha=0.8
)
plt.title('Operational Risk Map (Apriori): What Causes Severe Delays?')
plt.xlabel('Support (Frequency)')
plt.ylabel('Confidence (Likelihood)')
plt.legend(title='Lift')
plt.grid(True, alpha=0.3)
plt.show()
```

```
--- 4. Analysis: Predictors of Severe Delays ---
TOP RISK FACTORS (Apriori Results):
                                            antecedents           consequents  \
476             (Weather_Rain, Airline=JetBlue Airways)  (Delay_Class=Severe)
1102  (Tier_Name=Tier_1_HighRisk, Airline=JetBlue Ai…  (Delay_Class=Severe)
590           (Tier_Name=Tier_1_HighRisk, Weather_Rain)  (Delay_Class=Severe)
1769  (Weather_Rain, Tier_Name=Tier_4_MegaHub, Airli…  (Delay_Class=Severe)
149                           (Airline=JetBlue Airways)  (Delay_Class=Severe)
1116  (Tier_Name=Tier_4_MegaHub, Airline=JetBlue Air…  (Delay_Class=Severe)
1565  (Weather_Rain, Tier_Name=Tier_4_MegaHub, Airli…  (Delay_Class=Severe)
562       (Weather_Rain, Airline=United Air Lines Inc.)  (Delay_Class=Severe)
404      (Weather_Rain, Airline=American Airlines Inc.)  (Delay_Class=Severe)
608            (Weather_Rain, Tier_Name=Tier_4_MegaHub)  (Delay_Class=Severe)
1072  (Tier_Name=Tier_4_MegaHub, Airline=Frontier Ai…  (Delay_Class=Severe)
135                    (Airline=Frontier Airlines Inc.)  (Delay_Class=Severe)
1739  (Weather_Rain, Tier_Name=Tier_4_MegaHub, Airli…  (Delay_Class=Severe)
47                                       (Weather_Rain)  (Delay_Class=Severe)
180                          (Airline=Spirit Air Lines)  (Delay_Class=Severe)

       support  confidence      lift
476   0.001227    0.197461  2.922636
1102  0.001023    0.160729  2.378966
590   0.001258    0.157720  2.334428
1769  0.001330    0.149513  2.212953
149   0.005554    0.138531  2.050415
1116  0.003604    0.135888  2.011287
1565  0.001437    0.134728  1.994118
562   0.001690    0.133995  1.983273
404   0.002222    0.132350  1.958923
608   0.009004    0.127712  1.890275
1072  0.002206    0.124960  1.849542
135   0.003214    0.123799  1.832368
1739  0.001232    0.123740  1.831491
```

```
47     0.014010     0.115499   1.709512
180    0.004268     0.110333   1.633056
```



[7]:

```
# ==========================================
# CHUNK 6: EXPORTING ALL RULES
# ==========================================
print("--- 5. Exporting Full Rule Set ---")

# 1. Configure Pandas to show more data in the output window
pd.set_option('display.max_rows', 500)      # Show up to 500 rows
pd.set_option('display.max_colwidth', None) # Don't truncate the text

# 2. Sort ALL rules by Lift (Strength)
# We don't filter for 'Severe' here - this is EVERYTHING the algorithm found.
all_rules_sorted = rules.sort_values('lift', ascending=False)

# 3. Display the Top 50 to the screen
print("Top 50 Rules (All Consequences):")
cols = ['antecedents', 'consequents', 'support', 'confidence', 'lift']
print(all_rules_sorted[cols].head(50))

# 4. Save the COMPLETE list to CSV
```

```python
# This file will contain every single pattern found, including "Late",␣
 ↪"OnTime", etc.
filename = 'all_association_rules_full.csv'
all_rules_sorted.to_csv(filename, index=False)

print(f"\nSuccess! All {len(rules)} rules have been saved to '{filename}'.")
print("Open this file in Excel to explore the full dataset.")
```

```
--- 5. Exporting Full Rule Set ---
Top 50 Rules (All Consequences):
antecedents  \
1274                                    (Weather_Snow,
Tier_Name=Tier_3_Efficient)
1275                              (Airline=Alaska Airlines Inc.,
Weather_Freezing)
1272                                  (Weather_Snow, Airline=Alaska Airlines
Inc.)
1277                                (Weather_Freezing,
Tier_Name=Tier_3_Efficient)
1278
(Weather_Snow)
1271  (Airline=Alaska Airlines Inc., Weather_Freezing,
Tier_Name=Tier_3_Efficient)
1822                          (Airline=Hawaiian Airlines Inc.,
Delay_Class=OnTime)
1827                                    (Tier_Name=Tier_3_Efficient,
Weather_Wind)
650                                     (Tier_Name=Tier_3_Efficient,
Weather_Wind)
651                                               (Airline=Hawaiian Airlines
Inc.)
1821                (Delay_Class=OnTime, Tier_Name=Tier_3_Efficient,
Weather_Wind)
1828                                              (Airline=Hawaiian Airlines
Inc.)
1290                              (Airline=Alaska Airlines Inc.,
Weather_Freezing)
1287                                            (Weather_Snow,
Delay_Class=OnTime)
1285          (Airline=Alaska Airlines Inc., Delay_Class=OnTime,
Weather_Freezing)
1292
(Weather_Snow)
213
(Weather_Snow)
212                               (Airline=Alaska Airlines Inc.,
Weather_Freezing)
1257
```

```
(Weather_Snow)
1252                    (Weather_Rain, Tier_Name=Tier_4_MegaHub,
Weather_Freezing)
1463                            (Airline=Alaska Airlines Inc.,
Delay_Class=OnTime)
1462                                    (Weather_Snow,
Tier_Name=Tier_3_Efficient)
1279                                                (Airline=Alaska Airlines
Inc.)
1270                  (Weather_Snow, Weather_Freezing,
Tier_Name=Tier_3_Efficient)
1467                                                (Airline=Alaska Airlines
Inc.)
1458                (Weather_Snow, Delay_Class=OnTime,
Tier_Name=Tier_3_Efficient)
276                                                 (Airline=Alaska Airlines
Inc.)
273                                     (Weather_Snow,
Tier_Name=Tier_3_Efficient)
1408                                            (Weather_Snow,
Delay_Class=OnTime)
1411                                (Weather_Freezing,
Tier_Name=Tier_3_Efficient)
206
(Weather_Snow)
205                                               (Weather_Rain,
Weather_Freezing)
1255                                              (Weather_Rain,
Weather_Freezing)
1254                                      (Weather_Snow,
Tier_Name=Tier_4_MegaHub)
2143                  (Weather_Snow, Tier_Name=Tier_4_MegaHub,
Delay_Class=OnTime)
2148                               (Airline=Delta Air Lines Inc,
Weather_Freezing)
1452
(Weather_Snow)
1445              (Delay_Class=Severe, Tier_Name=Tier_4_MegaHub,
Weather_Freezing)
2178                                            (Weather_Snow,
Delay_Class=OnTime)
2169   (Tier_Name=Tier_4_MegaHub, Airline=Skywest Airlines Inc.,
Weather_Freezing)
248
(Weather_Snow)
247                                 (Weather_Freezing,
Tier_Name=Tier_3_Efficient)
1300                                      (Weather_Snow,
Tier_Name=Tier_4_MegaHub)
1305                               (Airline=Delta Air Lines Inc,
Weather_Freezing)
1325   (Tier_Name=Tier_4_MegaHub, Airline=Skywest Airlines Inc.,
Weather_Freezing)
1332
(Weather_Snow)
1407            (Delay_Class=OnTime, Weather_Freezing,
Tier_Name=Tier_3_Efficient)
1412
(Weather_Snow)
1261                          (Weather_Rain, Delay_Class=OnTime,
Weather_Freezing)
1266
(Weather_Snow)

consequents  \
1274                              (Airline=Alaska Airlines Inc.,
Weather_Freezing)
1275                                    (Weather_Snow,
Tier_Name=Tier_3_Efficient)
1272                                (Weather_Freezing,
Tier_Name=Tier_3_Efficient)
1277                                  (Weather_Snow, Airline=Alaska Airlines
Inc.)
1278  (Airline=Alaska Airlines Inc., Weather_Freezing,
Tier_Name=Tier_3_Efficient)
1271
(Weather_Snow)
1822                                    (Tier_Name=Tier_3_Efficient,
Weather_Wind)
1827                          (Airline=Hawaiian Airlines Inc.,
Delay_Class=OnTime)
650                                               (Airline=Hawaiian Airlines
Inc.)
651                                     (Tier_Name=Tier_3_Efficient,
Weather_Wind)
1821                                              (Airline=Hawaiian Airlines
Inc.)
1828                (Delay_Class=OnTime, Tier_Name=Tier_3_Efficient,
Weather_Wind)
1290                                            (Weather_Snow,
Delay_Class=OnTime)
1287                              (Airline=Alaska Airlines Inc.,
Weather_Freezing)
1285
(Weather_Snow)
1292          (Airline=Alaska Airlines Inc., Delay_Class=OnTime,
Weather_Freezing)
213                               (Airline=Alaska Airlines Inc.,
Weather_Freezing)
212
(Weather_Snow)
1257                    (Weather_Rain, Tier_Name=Tier_4_MegaHub,
Weather_Freezing)
1252
(Weather_Snow)
1463                                    (Weather_Snow,
Tier_Name=Tier_3_Efficient)
1462                            (Airline=Alaska Airlines Inc.,
Delay_Class=OnTime)
1279                  (Weather_Snow, Weather_Freezing,
Tier_Name=Tier_3_Efficient)
1270                                                (Airline=Alaska Airlines
Inc.)
1467                (Weather_Snow, Delay_Class=OnTime,
Tier_Name=Tier_3_Efficient)
1458                                                (Airline=Alaska Airlines
Inc.)
276                                     (Weather_Snow,
Tier_Name=Tier_3_Efficient)
273                                                 (Airline=Alaska Airlines
Inc.)
1408                                (Weather_Freezing,
Tier_Name=Tier_3_Efficient)
1411                                            (Weather_Snow,
Delay_Class=OnTime)
206                                               (Weather_Rain,
Weather_Freezing)
205
(Weather_Snow)
1255                                      (Weather_Snow,
Tier_Name=Tier_4_MegaHub)
1254                                              (Weather_Rain,
Weather_Freezing)
2143                               (Airline=Delta Air Lines Inc,
Weather_Freezing)
2148                  (Weather_Snow, Tier_Name=Tier_4_MegaHub,
Delay_Class=OnTime)
1452              (Delay_Class=Severe, Tier_Name=Tier_4_MegaHub,
Weather_Freezing)
1445
(Weather_Snow)
2178   (Tier_Name=Tier_4_MegaHub, Airline=Skywest Airlines Inc.,
Weather_Freezing)
2169                                            (Weather_Snow,
Delay_Class=OnTime)
248                                 (Weather_Freezing,
Tier_Name=Tier_3_Efficient)
247
(Weather_Snow)
1300                               (Airline=Delta Air Lines Inc,
Weather_Freezing)
1305                                      (Weather_Snow,
Tier_Name=Tier_4_MegaHub)
1325
(Weather_Snow)
1332   (Tier_Name=Tier_4_MegaHub, Airline=Skywest Airlines Inc.,
Weather_Freezing)
1407
(Weather_Snow)
1412            (Delay_Class=OnTime, Weather_Freezing,
Tier_Name=Tier_3_Efficient)
1261
(Weather_Snow)
1266                          (Weather_Rain, Delay_Class=OnTime,
Weather_Freezing)

       support  confidence        lift
1274  0.001146    0.522496  169.365624
1275  0.001146    0.371469  169.365624
1272  0.001146    0.794787  126.801290
1277  0.001146    0.182832  126.801290
1278  0.001146    0.048270   32.309135
1271  0.001146    0.767064   32.309135
1822  0.001185    0.124683   25.089429
1827  0.001185    0.238497   25.089429
650   0.001486    0.299003   25.047836
651   0.001486    0.124476   25.047836
1821  0.001185    0.293583   24.593835
1828  0.001185    0.099287   24.593835
1290  0.001088    0.352636   19.426743
1287  0.001088    0.059932   19.426743
1285  0.001088    0.424854   17.895057
1292  0.001088    0.045822   17.895057
213   0.001293    0.054463   17.654139
212   0.001293    0.419134   17.654139
1257  0.001156    0.048692   16.962458
1252  0.001156    0.402713   16.962458
1463  0.001059    0.036528   16.654181
1462  0.001059    0.482761   16.654181
1279  0.001146    0.032513   16.640208
1270  0.001146    0.586527   16.640208
1467  0.001059    0.030040   16.260813
```

```
1458  0.001059    0.573154   16.260813
276   0.001233    0.034985   15.950926
273   0.001233    0.562231   15.950926
1408  0.001639    0.090298   14.406243
1411  0.001639    0.261503   14.406243
206   0.001759    0.074111   14.073306
205   0.001759    0.334120   14.073306
1255  0.001156    0.219524   13.525471
1254  0.001156    0.071226   13.525471
2143  0.002758    0.228648   13.351842
2148  0.002758    0.161034   13.351842
1452  0.001224    0.051562   13.183320
1445  0.001224    0.312991   13.183320
2178  0.001930    0.106310   13.144722
2169  0.001930    0.238604   13.144722
248   0.001954    0.082297   13.129829
247   0.001954    0.311721   13.129829
1300  0.003586    0.220929   12.901062
1305  0.003586    0.209390   12.901062
1325  0.002476    0.306147   12.895067
1332  0.002476    0.104291   12.895067
1407  0.001639    0.304987   12.846226
1412  0.001639    0.069039   12.846226
1261  0.001093    0.299369   12.609555
1266  0.001093    0.046049   12.609555

Success! All 2238 rules have been saved to 'all_association_rules_full.csv'.
Open this file in Excel to explore the full dataset.
```