One potential use of the data pipeline is curated flight information for travelers. The following is a sample output showing the yearly departure count and average total delay (departure and arrival) categorized by airline at each airport.

import pandas as pd import numpy as numpy import matplotlib.pyplot as plt import warnings import getpass warnings.filterwarnings('ignore') In [3]: #connect to offline database

conn=mysql.connect(host='localhost',

port=int(3306), user='root',

Alaska Airlines Inc.

JetBlue Airways

In [1]: #import necessary libraries import pymysql as mysql

Enter password: .....

passwd=getpass.getpass('Enter password:'), db='ads507airlines')

In [9]: #check connection check = pd.read sql ("SELECT \* from airlines all2018""",conn)

id carrier

Out[9]: 0 1 9E

AS

В6

2 3

3

name Endeavor Air Inc. American Airlines Inc. AΑ

yearlySEA= pd.read\_sql ("""SELECT \* from yearly\_delay2018 WHERE origin='SEA'""", conn) yearlySAN= pd.read\_sql ("""SELECT \* from yearly\_delay2018 WHERE origin='SAN'""", conn)

UA

ΕV

AA

Republic Airline

Mesa Airlines Inc.

plt.title("Yearly Departures from JFK by Airline (2018)")

American Airlines Inc. -

Spirit Air Lines

SkyWest Airlines Inc.

Delta Air Lines Inc.

Airline

Frontier Airlines Inc.

Endeavor Air Inc.

Alaska Airlines Inc.

Southwest Airlines Co.

Envoy Air

JetBlue Airways

25

United Air Lines Inc. -

JFK (2018)

plt.xticks(rotation=90)

plt.xlabel("Airline") plt.ylabel("Flight Count")

In [60]: x1=yearlyJFK['name']

plt.bar(x1,y1)

plt.twinx()

plt.show()

ExpressJet Airlines LLC d/b/a aha! -

y1=yearlyJFK['yearly flight count'] fig = plt.figure(figsize =(12, 7))

y1 delay=yearlyJFK['total delay'] plt.scatter(x1,y1\_delay,color='black') plt.ylabel("Average Delay (min)")

IAH

IAH

IAH

IAH

name carrier origin yearly\_flight\_count total\_delay

13.2871

2.8621

3.6284

17.3182

66215

38142

31979

13132

8627

Delta Air Lines Inc. 5 DL ExpressJet Airlines LLC d/b/a aha! 7 F9 Frontier Airlines Inc. 8 G4 Allegiant Air Hawaiian Airlines Inc. 9 8 HA 10 MQ Envoy Air Spirit Air Lines **10** 11 NK **11** 12 PSA Airlines Inc. OH SkyWest Airlines Inc. **12** 13 00 Horizon Air United Air Lines Inc. **14** 15 UA Virgin America 15 16 VX **16** 17 WN Southwest Airlines Co. Mesa Airlines Inc. **18** 19 ΥX Republic Airline In [35]: #partition yearly flight count and average total delays by each airport and save as pandas dataframe yearlyIAH= pd.read\_sql ("""SELECT \* from yearly\_delay2018 WHERE origin='IAH'""", conn) yearlyJFK= pd.read sql ("""SELECT \* from yearly delay2018 WHERE origin='JFK'""", conn)

United Air Lines Inc.

Mesa Airlines Inc.

American Airlines Inc.

Republic Airline

Yearly Departures by Airline

1 ExpressJet Airlines LLC d/b/a aha!

In [76]: x=yearlyIAH['name'] y=yearlyIAH['yearly\_flight\_count'] fig = plt.figure(figsize =(12, 7))

plt.rcParams.update({'font.size': 14})

## plt.title("Yearly Departures from IAH by Airline (2018)") plt.xlabel("Airline") plt.ylabel("Flight Count")

plt.bar(x,y)

plt.xticks(rotation=90)

IAH (2018)

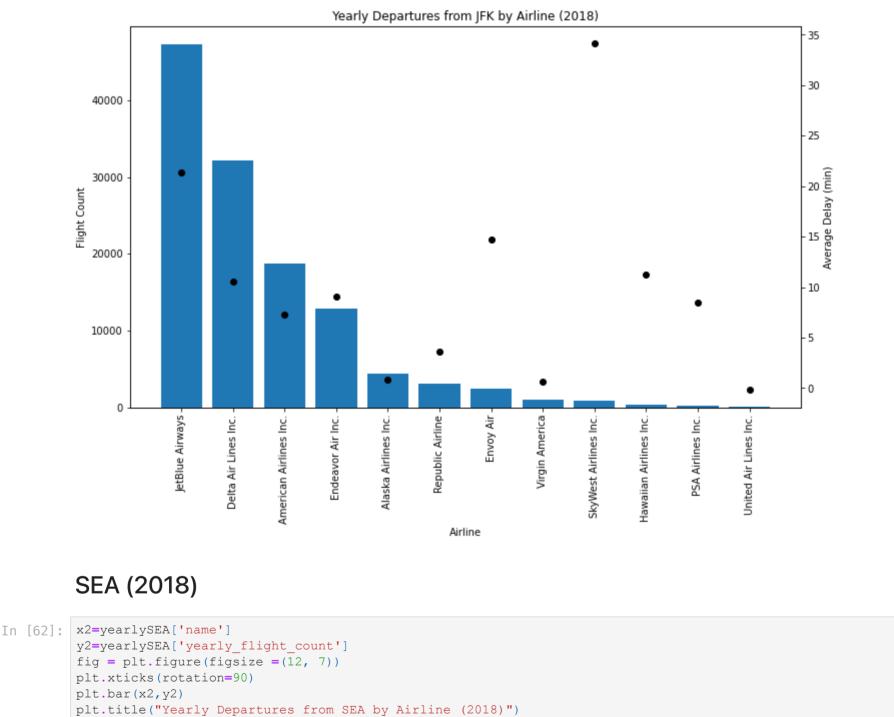
yearlyIAH.head(5)

Out[35]:

4

plt.twinx()

y\_delay=yearlyIAH['total\_delay'] plt.scatter(x,y\_delay,color='orange') plt.ylabel("Average Delay (min)") plt.show() Yearly Departures from IAH by Airline (2018) 20.0 60000 17.5 50000 15.0 0.01 Average Delay (min) Flight Count 40000 30000 20000 7.5 5.0 10000 2.5



## 70000

plt.twinx()

plt.show()

plt.xlabel("Airline") plt.ylabel("Flight Count")

60000 50000

y2\_delay=yearlySEA['total\_delay'] plt.scatter(x2, y2\_delay, color='black') plt.ylabel("Average Delay (min)")

51 Average Delay (min) Flight Count 40000 30000 10 20000 10000 5 Alaska Airlines Inc. etBlue Airways Spirit Air Lines Virgin America Delta Air Lines Inc Southwest Airlines Co. United Air Lines Inc SkyWest Airlines Inc American Airlines Inc Hawaiian Airlines Inc Frontier Airlines Inc Airline SAN (2018) In [63]: x3=yearlySAN['name'] y3=yearlySAN['yearly\_flight\_count'] fig = plt.figure(figsize =(12, 7)) plt.xticks(rotation=90) plt.bar(x3, y3)plt.title("Yearly Departures from SAN by Airline (2018)") plt.xlabel("Airline")

Yearly Departures from SEA by Airline (2018)

## plt.ylabel("Average Delay (min)") plt.show()

plt.twinx()

plt.ylabel("Flight Count")

y3\_delay=yearlySAN['total\_delay'] plt.scatter(x3, y3\_delay, color='black')

Yearly Departures from SAN by Airline (2018) 40000 35000

40 30000 No 85 Average Delay (min) 25000 Flight Count 20000 15000 10000 10 5000 Spirit Air Lines JetBlue Airways Virgin America Allegiant Air Horizon Air Southwest Airlines Co. SkyWest Airlines Inc. Hawaiian Airlines Inc. United Air Lines Inc. Alaska Airlines Inc. American Airlines Inc. Delta Air Lines Inc. Frontier Airlines Inc.

Airline