

In [18]:

```
import pyspark
from pyspark.sql import SparkSession
from pyspark.sql.functions import concat, col, lit
from sedona.register import SedonaRegistrator
from sedona.utils import SedonaKryoRegistrator, KryoSerializer

import geopandas as gpd
spark = SparkSession. \
    builder. \
    appName('Milestone 2 Data Prep'). \
    master("spark://nwalton-ubuntu:7077"). \
    config("spark.serializer", KryoSerializer.getName). \
    config("spark.kryo.registrator", SedonaKryoRegistrator.getName). \
    config('spark.jars.packages',
           'org.apache.sedona:sedona-python-adapter-3.0_2.12:1.0.1-incubating,'
           'org.datasyslab:geotools-wrapper:geotools-24.1'). \
    getOrCreate()

SedonaRegistrator.registerAll(spark)
```

```
21/11/14 22:45:37 WARN UDTRegistration: Cannot register UDT for org.locationtech.jts.geom.
Geometry, which is already registered.
21/11/14 22:45:37 WARN UDTRegistration: Cannot register UDT for org.locationtech.jts.inde
x.SpatialIndex, which is already registered.
21/11/14 22:45:37 WARN SimpleFunctionRegistry: The function st_pointfromtext replaced a pr
eviously registered function.
21/11/14 22:45:37 WARN SimpleFunctionRegistry: The function st_polygonfromtext replaced a
previously registered function.
21/11/14 22:45:37 WARN SimpleFunctionRegistry: The function st_linestringfromtext replaced
a previously registered function.
21/11/14 22:45:37 WARN SimpleFunctionRegistry: The function st_geomfromtext replaced a pr
eviously registered function.
21/11/14 22:45:37 WARN SimpleFunctionRegistry: The function st_geomfromwkt replaced a prev
iously registered function.
21/11/14 22:45:37 WARN SimpleFunctionRegistry: The function st_geomfromwkb replaced a prev
iously registered function.
21/11/14 22:45:37 WARN SimpleFunctionRegistry: The function st_geomfromgeojson replaced a
previously registered function.
21/11/14 22:45:37 WARN SimpleFunctionRegistry: The function st_point replaced a previously
registered function.
21/11/14 22:45:37 WARN SimpleFunctionRegistry: The function st_polygonfromenvelope replace
d a previously registered function.
21/11/14 22:45:37 WARN SimpleFunctionRegistry: The function st_contains replaced a previou
sly registered function.
21/11/14 22:45:37 WARN SimpleFunctionRegistry: The function st_intersects replaced a previ
ously registered function.
21/11/14 22:45:37 WARN SimpleFunctionRegistry: The function st_within replaced a previousl
y registered function.
21/11/14 22:45:37 WARN SimpleFunctionRegistry: The function st_distance replaced a previou
sly registered function.
21/11/14 22:45:37 WARN SimpleFunctionRegistry: The function st_convexhull replaced a previ
ously registered function.
True
```

Out[18]:

```
21/11/14 22:45:37 WARN SimpleFunctionRegistry: The function st_npoints replaced a previous
ly registered function.
21/11/14 22:45:37 WARN SimpleFunctionRegistry: The function st_buffer replaced a previousl
y registered function.
21/11/14 22:45:37 WARN SimpleFunctionRegistry: The function st_envelope replaced a previou
sly registered function.
21/11/14 22:45:37 WARN SimpleFunctionRegistry: The function st_length replaced a previousl
y registered function.
21/11/14 22:45:37 WARN SimpleFunctionRegistry: The function st_area replaced a previously
```

registered function.  
21/11/14 22:45:37 WARN SimpleFunctionRegistry: The function st\_centroid replaced a previously registered function.  
21/11/14 22:45:37 WARN SimpleFunctionRegistry: The function st\_transform replaced a previously registered function.  
21/11/14 22:45:37 WARN SimpleFunctionRegistry: The function st\_intersection replaced a previously registered function.  
21/11/14 22:45:37 WARN SimpleFunctionRegistry: The function st\_isvalid replaced a previously registered function.  
21/11/14 22:45:37 WARN SimpleFunctionRegistry: The function st\_precisionreduce replaced a previously registered function.  
21/11/14 22:45:37 WARN SimpleFunctionRegistry: The function st\_equals replaced a previously registered function.  
21/11/14 22:45:37 WARN SimpleFunctionRegistry: The function st\_touches replaced a previously registered function.  
21/11/14 22:45:37 WARN SimpleFunctionRegistry: The function st\_overlaps replaced a previously registered function.  
21/11/14 22:45:37 WARN SimpleFunctionRegistry: The function st\_crosses replaced a previously registered function.  
21/11/14 22:45:37 WARN SimpleFunctionRegistry: The function st\_issimple replaced a previously registered function.  
21/11/14 22:45:37 WARN SimpleFunctionRegistry: The function st\_makevalid replaced a previously registered function.  
21/11/14 22:45:37 WARN SimpleFunctionRegistry: The function st\_simplifypreservetopology replaced a previously registered function.  
21/11/14 22:45:37 WARN SimpleFunctionRegistry: The function st\_astext replaced a previously registered function.  
21/11/14 22:45:37 WARN SimpleFunctionRegistry: The function st\_asgeojson replaced a previously registered function.  
21/11/14 22:45:37 WARN SimpleFunctionRegistry: The function st\_geometrytype replaced a previously registered function.  
21/11/14 22:45:37 WARN SimpleFunctionRegistry: The function st\_numgeometries replaced a previously registered function.  
21/11/14 22:45:37 WARN SimpleFunctionRegistry: The function st\_linemerge replaced a previously registered function.  
21/11/14 22:45:37 WARN SimpleFunctionRegistry: The function st\_azimuth replaced a previously registered function.  
21/11/14 22:45:37 WARN SimpleFunctionRegistry: The function st\_x replaced a previously registered function.  
21/11/14 22:45:37 WARN SimpleFunctionRegistry: The function st\_y replaced a previously registered function.  
21/11/14 22:45:37 WARN SimpleFunctionRegistry: The function st\_startpoint replaced a previously registered function.  
21/11/14 22:45:37 WARN SimpleFunctionRegistry: The function st\_boundary replaced a previously registered function.  
21/11/14 22:45:37 WARN SimpleFunctionRegistry: The function st\_minimumboundingradius replaced a previously registered function.  
21/11/14 22:45:37 WARN SimpleFunctionRegistry: The function st\_minimumboundingcircle replaced a previously registered function.  
21/11/14 22:45:37 WARN SimpleFunctionRegistry: The function st\_endpoint replaced a previously registered function.  
21/11/14 22:45:37 WARN SimpleFunctionRegistry: The function st\_exteriorring replaced a previously registered function.  
21/11/14 22:45:37 WARN SimpleFunctionRegistry: The function st\_geometryn replaced a previously registered function.  
21/11/14 22:45:37 WARN SimpleFunctionRegistry: The function st\_interiorringn replaced a previously registered function.  
21/11/14 22:45:37 WARN SimpleFunctionRegistry: The function st\_dump replaced a previously registered function.  
21/11/14 22:45:37 WARN SimpleFunctionRegistry: The function st\_dumppoints replaced a previously registered function.  
21/11/14 22:45:37 WARN SimpleFunctionRegistry: The function st\_isclosed replaced a previously registered function.  
21/11/14 22:45:37 WARN SimpleFunctionRegistry: The function st\_numinteriorrings replaced a previously registered function.  
21/11/14 22:45:37 WARN SimpleFunctionRegistry: The function st\_addpoint replaced a previously

sly registered function.  
 21/11/14 22:45:37 WARN SimpleFunctionRegistry: The function st\_removepoint replaced a previously registered function.  
 21/11/14 22:45:37 WARN SimpleFunctionRegistry: The function st\_isring replaced a previously registered function.  
 21/11/14 22:45:37 WARN SimpleFunctionRegistry: The function st\_flipcoordinates replaced a previously registered function.  
 21/11/14 22:45:37 WARN SimpleFunctionRegistry: The function st\_linesubstring replaced a previously registered function.  
 21/11/14 22:45:37 WARN SimpleFunctionRegistry: The function st\_lineinterpolatepoint replaced a previously registered function.  
 21/11/14 22:45:37 WARN SimpleFunctionRegistry: The function st\_union\_aggr replaced a previously registered function.  
 21/11/14 22:45:37 WARN SimpleFunctionRegistry: The function st\_envelope\_aggr replaced a previously registered function.  
 21/11/14 22:45:37 WARN SimpleFunctionRegistry: The function st\_intersection\_aggr replaced a previously registered function.

## Data Prep

In [3]:

```
#Airport DF
airports_dat = spark.read.format("csv").\
    option("header", "true").\
    option("delimiter", ",").\
    load("./Data/airports.dat").\
    toDF("Airport_ID", "Name", "City", "Country", "IATA", "ICAO",
        "Latitude", "Longitude", "Altitude", "Timezone", "DST", "Tz Database time zone",
        "Type", "Source")

airports_dat.show()
airports_dat.printSchema()
```

```
+-----+-----+-----+-----+-----+-----+-----+-----+
+-----+-----+-----+-----+-----+-----+-----+-----+
+-+
|Airport_ID|          Name|          City|          Country|IATA|ICAO|          L
atititude|          Longitude|Altitude|Timezone|DST|Tz Database time zone|  Type|          Sourc
e|
+-----+-----+-----+-----+-----+-----+-----+-----+
+-----+-----+-----+-----+-----+-----+-----+-----+
+-+
|          2|    Madang Airport|    Madang|Papua New Guinea| MAG|AYMD|    -5.2070
7988739|    145.789001465|    20|    10|  U| Pacific/Port_Moresby|airport|OurAirport
s|
|          3|Mount Hagen Kagam...|    Mount Hagen|Papua New Guinea| HGU|AYMH|-5.82678985
5957031|  144.29600524902344|    5388|    10|  U| Pacific/Port_Moresby|airport|OurAirport
s|
|          4|    Nadzab Airport|    Nadzab|Papua New Guinea| LAE|AYNZ|    -
6.569803|    146.725977|    239|    10|  U| Pacific/Port_Moresby|airport|OurAirpor
ts|
|          5|Port Moresby Jack...|    Port Moresby|Papua New Guinea| POM|AYPY|-9.44338035
5834961|  147.22000122070312|    146|    10|  U| Pacific/Port_Moresby|airport|OurAirport
s|
|          6|Wewak Internation...|    Wewak|Papua New Guinea| WWK|AYWK|    -3.5838
3011818|    143.669006348|    19|    10|  U| Pacific/Port_Moresby|airport|OurAirport
s|
|          7|  Narsarsuaq Airport|  Narssarssuaq|    Greenland| UAK|BGBW|    61.160
4995728|   -45.4259986877|    112|   -3|  E|    America/Godthab|airport|OurAirport
s|
|          8|Godthaab / Nuuk A...|    Godthaab|    Greenland| GOH|BGGH|    64.1
9090271|   -51.6781005859|    283|   -3|  E|    America/Godthab|airport|OurAirport
s|
|          9|Kangerlussuaq Air...|    Sondrestrom|    Greenland| SFJ|BGSF|    67.012
```



```
state_boundary_dat.show()
```

```
+-----+-----+
|      s_name|      s_bound|
+-----+-----+
|      Alaska|POLYGON((-141.020...|
|      Alabama|POLYGON((-88.1955...|
|      Arkansas|POLYGON((-94.0416...|
|      Arizona|POLYGON((-112.598...|
|      California|POLYGON((-124.400...|
|      Colorado|POLYGON((-109.044...|
|      Connecticut|POLYGON((-73.4875...|
|      Delaware|POLYGON((-75.7919...|
|      Florida|POLYGON((-87.6050...|
|      Georgia|POLYGON((-85.6082...|
|      Hawaii|POLYGON((-154.628...|
|      Iowa|POLYGON((-95.7623...|
|      Idaho|POLYGON((-117.031...|
|      Illinois|POLYGON((-90.6290...|
|      Indiana|POLYGON((-87.5253...|
|      Kansas|POLYGON((-102.050...|
|      Kentucky|POLYGON((-89.5372...|
|      Louisiana|POLYGON((-94.0430...|
|      Massachusetts|POLYGON((-72.7789...|
|      Maryland|POLYGON((-79.4778...|
+-----+-----+
```

only showing top 20 rows

In [5]:

```
airlines_dat = spark.read.format("csv").\
option("header", "true").\
option("delimiter", ",").\
load("Data/airlines.dat").\
toDF("Airline_ID", "Name", "Alias", "IATA", "ICAO", "Callsign",
     "Country", "Active")

airlines_dat.show()
airlines_dat.printSchema()
```

```
+-----+-----+-----+-----+-----+-----+-----+-----+
|Airline_ID|      Name|Alias|IATA|ICAO|      Callsign|      Country|Active|
+-----+-----+-----+-----+-----+-----+-----+-----+
|      1|Private flight|\N|-|N/A|      null|      null|Y|
|      2|135 Airways|\N|null|GNL|      GENERAL|United States|N|
|      3|1Time Airline|\N|1T|RXN|      NEXTIME|South Africa|Y|
|      4|2 Sqn No 1 Elemen...\N|null|WYT|      null|United Kingdom|N|
|      5|213 Flight Unit|\N|null|TFU|      null|Russia|N|
|      6|223 Flight Unit S...\N|null|CHD|CHKALOVSK-AVIA|Russia|N|
|      7|224th Flight Unit|\N|null|TTF|      CARGO UNIT|Russia|N|
|      8|247 Jet Ltd|\N|null|TWF|      CLOUD RUNNER|United Kingdom|N|
|      9|3D Aviation|\N|null|SEC|      SECUREX|United States|N|
|     10|40-Mile Air|\N|Q5|MLA|      MILE-AIR|United States|Y|
|     11|4D Air|\N|null|QRT|      QUARTET|Thailand|N|
|     12|611897 Alberta Li...\N|null|THD|      DONUT|Canada|N|
|     13|Ansett Australia|\N|AN|AAA|      ANSETT|Australia|Y|
|     14|Abacus International|\N|1B|null|      null|Singapore|Y|
|     15|Abelag Aviation|\N|W9|AAB|      ABG|Belgium|N|
|     16|Army Air Corps|\N|null|AAC|      ARMYAIR|United Kingdom|N|
|     17|Aero Aviation Cen...\N|null|AAD|      SUNRISE|Canada|N|
|     18|Aero Servicios Ej...\N|null|SII|      ASEISA|Mexico|N|
|     19|Aero Biniza|\N|null|BZS|      BINIZA|Mexico|N|
|     20|Aero Albatros|\N|null|ABM|ALBATROS ESPANA|Spain|N|
+-----+-----+-----+-----+-----+-----+-----+-----+
```

only showing top 20 rows

```

root
|-- Airline_ID: string (nullable = true)
|-- Name: string (nullable = true)
|-- Alias: string (nullable = true)
|-- IATA: string (nullable = true)
|-- ICAO: string (nullable = true)
|-- Callsign: string (nullable = true)
|-- Country: string (nullable = true)
|-- Active: string (nullable = true)

```

In [6]:

```

routes_dat = spark.read.format("csv").\
option("header", "true").\
option("delimiter", ",").\
load("Data/routes.dat").\
toDF("Airline", "Airline_ID", "Source_Airport", "Source_Airport_ID",
     "Destination_Airport", "Destination_Airport_ID", "Codeshare",
     "Stops", "Equipment")

routes_dat.show()
routes_dat.printSchema()

```

```

+-----+-----+-----+-----+-----+-----+-----+-----+
|Airline|Airline_ID|Source_Airport|Source_Airport_ID|Destination_Airport|Destination_Airpo
rt_ID|Codeshare|Stops|Equipment|
+-----+-----+-----+-----+-----+-----+-----+-----+
|      2B|      410|      ASF|      2966|      KZN|
2990|      null|      0|      CR2|
|      2B|      410|      ASF|      2966|      MRV|
2962|      null|      0|      CR2|
|      2B|      410|      CEK|      2968|      KZN|
2990|      null|      0|      CR2|
|      2B|      410|      CEK|      2968|      OVB|
4078|      null|      0|      CR2|
|      2B|      410|      DME|      4029|      KZN|
2990|      null|      0|      CR2|
|      2B|      410|      DME|      4029|      NBC|
6969|      null|      0|      CR2|
|      2B|      410|      DME|      4029|      TGK|
|      \N|      null|      0|      CR2|
|      2B|      410|      DME|      4029|      UUA|
6160|      null|      0|      CR2|
|      2B|      410|      EGO|      6156|      KGD|
2952|      null|      0|      CR2|
|      2B|      410|      EGO|      6156|      KZN|
2990|      null|      0|      CR2|
|      2B|      410|      GYD|      2922|      NBC|
6969|      null|      0|      CR2|
|      2B|      410|      KGD|      2952|      EGO|
6156|      null|      0|      CR2|
|      2B|      410|      KZN|      2990|      AER|
2965|      null|      0|      CR2|
|      2B|      410|      KZN|      2990|      ASF|
2966|      null|      0|      CR2|
|      2B|      410|      KZN|      2990|      CEK|
2968|      null|      0|      CR2|
|      2B|      410|      KZN|      2990|      DME|
4029|      null|      0|      CR2|
|      2B|      410|      KZN|      2990|      EGO|
6156|      null|      0|      CR2|
|      2B|      410|      KZN|      2990|      LED|
2948|      null|      0|      CR2|

```

```

|      2B|      410|      KZN|      2990|      SVX|
2975|      null|      0|      CR2|
|      2B|      410|      LED|      2948|      KZN|
2990|      null|      0|      CR2|
+-----+-----+-----+-----+-----+-----+-----+
-----+-----+-----+-----+
only showing top 20 rows

```

```

root
|-- Airline: string (nullable = true)
|-- Airline_ID: string (nullable = true)
|-- Source_Airport: string (nullable = true)
|-- Source_Airport_ID: string (nullable = true)
|-- Destination_Airport: string (nullable = true)
|-- Destination_Airport_ID: string (nullable = true)
|-- Codeshare: string (nullable = true)
|-- Stops: string (nullable = true)
|-- Equipment: string (nullable = true)

```

## Milestone 3 Tasks

### Airport and Airline search

#### List of airports operating in country x

```

In [7]: #filter by desired country, united states for example
United_States_Airports = airports_dat.filter(airports_dat["Country"] == "United States")
United_States_Airports.show()

```

```

+-----+-----+-----+-----+-----+-----+-----+
--+-----+-----+-----+-----+-----+-----+-----+
|Airport_ID|      Name|      City|      Country|IATA|ICAO|      Latitu
de|      Longitude|Altitude|Timezone|DST|Tz Database time zone|      Type|      Source|
+-----+-----+-----+-----+-----+-----+-----+
--+-----+-----+-----+-----+-----+-----+-----+
|      3411|Barter Island LRR...|      Barter Island|United States| BTI|PABA|      70.13400268
55|      -143.582000732|      2|      -9|      A|      America/Anchorage|airport|OurAirports|
|      3412|Wainwright Air St...|      Fort Wainwright|United States|      \N|PAWT|      70.613403
32|      -159.86000006|      35|      -9|      A|      America/Anchorage|airport|OurAirports|
|      3413|Cape Lisburne LRR...|      Cape Lisburne|United States|      LUR|PALU|      68.875099
18|      -166.11000006|      16|      -9|      A|      America/Anchorage|airport|OurAirports|
|      3414|Point Lay LRRS Ai...|      Point Lay|United States|      PIZ|PPIZ|      69.732902
53|      -163.00500049|      22|      -9|      A|      America/Anchorage|airport|OurAirports|
|      3415|Hilo Internationa...|      Hilo|United States|      ITO|PHTO|19.7213993072509
77|-155.04800415039062|      38|      -10|      N|      Pacific/Honolulu|airport|OurAirports|
|      3416|Orlando Executive...|      Orlando|United States|      ORL|KORL|      28.54
55|      -81.332901|      113|      -5|      A|      America/New_York|airport|OurAirports|
|      3417|      Bettles Airport|      Bettles|United States|      BTT|PABT|      66.913902
28|      -151.529007|      647|      -9|      A|      America/Anchorage|airport|OurAirports|
|      3418|      Clear Airport|      Clear Mews|United States|      \N|PACL|      64.3012
01|      -149.119995|      552|      -9|      A|      America/Anchorage|airport|OurAirports|
|      3419|Indian Mountain L...|      Indian Mountains|United States|      UTO|PAIM|      65.992797
85|      -153.7039948|      1273|      -9|      A|      America/Anchorage|airport|OurAirports|
|      3420|      Fort Yukon Airport|      Fort Yukon|United States|      FYU|PFYU|      66.571502685546
88|      -145.25|      433|      -9|      A|      America/Anchorage|airport|OurAirports|
|      3421|Sparrevohn LRRS A...|      Sparrevohn|United States|      SVW|PASV|      61.097400
67|      -155.5740051|      1585|      -9|      A|      America/Anchorage|airport|OurAirports|
|      3422|Bryant Army Heliport|      Fort Richardson|United States|      FRN|PAFR|      61.266399
38|      -149.6529999|      378|      -9|      A|      America/Anchorage|airport|OurAirports|
|      3423|Tatalina LRRS Air...|      Tatalina|United States|      TLJ|PATL|62.8944015502999

```



```

96|      -155.977005005|      964|      -9|  A|      America/Anchorage|airport|OurAirports|
|      3424|Cape Romanzof LRR...|      CapeRomanzof|United States| CZF|PACZ|      61.780300
14|      -166.0390015|      464|      -9|  A|      America/Anchorage|airport|OurAirports|
|      3425|Laurence G Hansco...|      Bedford|United States| BED|KBED|      42.470001
22|      -71.28900146|      133|      -5|  A|      America/New_York|airport|OurAirports|
|      3426|St Paul Island Ai...| St. Paul Island|United States| SNP|PASN|57.1673011779785
16|-170.22000122070312|      63|      -9|  A|      America/Anchorage|airport|OurAirports|
|      3427|Cape Newenham LRR...| Cape Newenham|United States| EHM|PAEH|58.6464004516999
96|      -162.06300354|      541|      -9|  A|      America/Anchorage|airport|OurAirports|
|      3428|      St George Airport|      Point Barrow|United States| STG|PAPB|56.5783004761000
04|      -169.662002563|      125|      -9|  A|      America/Anchorage|airport|OurAirports|
|      3429|      Iliamna Airport|      Iliamna|United States| ILI|PAIL|      59.754398
35|      -154.9109955|      192|      -9|  A|      America/Anchorage|airport|OurAirports|
|      3430|      Platinum Airport|      Port Moller|United States| PTU|PAPM| 59.011398315429
69|-161.82000732421875|      15|      -9|  A|      America/Anchorage|airport|OurAirports|
+-----+-----+-----+-----+-----+-----+-----+-----+
--+-+-----+-----+-----+-----+-----+-----+-----+
only showing top 20 rows

```

## List of airlines having x stops

In [8]:

```

X_Stops = "1"
#Filter routes_dat by routes with only one stop
X_Stop_Flight_Routes = routes_dat.filter(routes_dat["Stops"] == X_Stops)

# Create a list of airline_IDs from routes with one stop
Airline_ID_List = [row[0] for row in X_Stop_Flight_Routes.select("Airline_ID").distinct()]

# Filter airline_dat by airline_IDs from the list above
X_Stop_Airlines = airlines_dat.filter(airlines_dat["Airline_ID"].isin(Airline_ID_List))
X_Stop_Airlines.show()

```

```

+-----+-----+-----+-----+-----+-----+-----+-----+
+-----+
|Airline_ID|      Name|      Alias|IATA|ICAO|      Callsign|      Country
|Active|
+-----+-----+-----+-----+-----+-----+-----+
+-----+
|      330|      Air Canada|      \N| AC| ACA| AIR CANADA|      Canada
|      Y|
|      1316|      AirTran Airways|      \N| FL| TRS|      CITRUS|United States
|      Y|
|      1623|      Canadian North|      \N| 5T| MPE|      EMPRESS|      Canada
|      Y|
|      1936|      Cubana de Aviación|      \N| CU| CUB|      CUBANA|      Cuba
|      Y|
|      4319|Scandinavian Airl...|SAS Scandinavian ...| SK| SAS|SCANDINAVIAN|      Sweden
|      Y|
|      4547|      Southwest Airlines|      \N| WN| SWA|      SOUTHWEST|United States
|      Y|
+-----+-----+-----+-----+-----+-----+-----+
+-----+

```

## List of airlines operating with code share

In [9]:

```

# Filter routes_dat by routes with codeshare
Codeshare_Routes = routes_dat.filter(routes_dat["Codeshare"] == "Y")

# Create a list of airline_IDs marked Y for codeshare
Codeshare_List = [row[0] for row in Codeshare_Routes.select("Airline_ID").distinct().collect()]

```



```
# Filter airline_dat by airline_IDs from the above list
Airlines_With_Codeshare = airlines_dat.filter(airlines_dat["Airline_ID"].isin(Codeshare_Li
Airlines_With_Codeshare.show()
```

```
+-----+-----+-----+-----+-----+-----+-----+-----+
-----+-----+
|Airline_ID|          Name|          Alias|IATA|ICAO|          Callsign|
Country|Active|
+-----+-----+-----+-----+-----+-----+-----+-----+
-----+-----+
|          24| American Airlines|          \N| AA| AAL|          AMERICAN|          United
States|          Y|
|          28| Asiana Airlines|          \N| OZ| AAR|          ASIANA|Republic o
f Korea|          Y|
|          90| Air Europa|          \N| UX| AEA|          EUROPA|
Spain|          Y|
|          96| Aegean Airlines|          \N| A3| AEE|          AEGEAN|
Greece|          Y|
|          116| Air Italy|          \N| I9| AEY|          AIR ITALY|
Italy|          Y|
|          130|Aeroflot Russian ...|          \N| SU| AFL|          AEROFLOT|
Russia|          Y|
|          137| Air France|          \N| AF| AFR|          AIRFRANS|
France|          Y|
|          197| Azerbaijan Airlines|          \N| J2| AHY|          AZAL|          Aze
rbaijan|          Y|
|          214| Air Berlin|          \N| AB| BER|          AIR BERLIN|
Germany|          Y|
|          218| Air India Limited|          \N| AI| AIC|          AIRINDIA|
India|          Y|
|          225| Air Tahiti Nui|          \N| TN| THT|TAHITI AIRLINES|
France|          Y|
|          240| Air One|          \N| AP| ADH|          HERON|
Italy|          Y|
|          241| Air Sahara|          \N| S2| RSH|          SAHARA|
India|          Y|
|          242| Air Malta|          \N| KM| AMC|          AIR MALTA|
Malta|          Y|
|          316| Air Macau|          \N| NX| AMU|          AIR MACAO|
Macao|          Y|
|          319| Air Seychelles|          \N| HM| SEY|          SEYCHELLES|          Sey
chelles|          Y|
|          321| AeroMéxico|          \N| AM| AMX|          AEROMEXICO|
Mexico|          Y|
|          324| All Nippon Airways|ANA All Nippon Ai...| NH| ANA|          ALL NIPPON|
Japan|          Y|
|          330| Air Canada|          \N| AC| ACA|          AIR CANADA|
Canada|          Y|
|          333| Air Baltic|          \N| BT| BTI|          AIRBALTIC|
Latvia|          Y|
+-----+-----+-----+-----+-----+-----+-----+-----+
-----+-----+
only showing top 20 rows
```

## List of active airlines in the United States

```
In [10]: US_Airlines = airlines_dat.filter(airlines_dat["Country"] == "United States")
Active_US_Airlines = US_Airlines.filter(US_Airlines["Active"] == "Y")

Active_US_Airlines.show()
```

Airline_ID	Name	Alias	IATA	ICAO	Callsign	Country	Active
10	40-Mile Air	\N	Q5	MLA	MILE-AIR	United States	Y
22	Aloha Airlines	\N	AQ	AAH	ALOHA	United States	Y
24	American Airlines	\N	AA	AAL	AMERICAN	United States	Y
35	Allegiant Air	\N	G4	AAY	ALLEGIANT	United States	Y
109	Alaska Central Ex...	\N	K0	AER	ACE AIR	United States	Y
149	Air Cargo Carriers	\N	2Q	SNC	NIGHT CARGO	United States	Y
210	Airlift Internati...	\N	null	AIR	AIRLIFT	United States	Y
281	America West Airl...	\N	HP	AWE	CACTUS	United States	Y
282	Air Wisconsin	\N	ZW	AWI	AIR WISCONSIN	United States	Y
287	Allegheny Commute...	\N	null	ALO	ALLEGHENY	United States	Y
295	Air Sunshine	\N	null	RSI	AIR SUNSHINE	United States	Y
315	ATA Airlines	\N	null	AMT	AMTRAN	United States	Y
397	Arrow Air	\N	JW	APW	BIG A	United States	Y
452	Atlantic Southeas...	\N	EV	ASQ	ACEY	United States	Y
659	American Eagle Ai...	\N	MQ	EGF	EAGLE FLIGHT	United States	Y
792	Access Air	\N	ZA	CYD	CYCLONE	United States	Y
882	Air Florida	\N	QH	FLZ	AIR FLORIDA	United States	Y
928	Atlas Air	\N	5Y	GTI	GIANT	United States	Y
1316	AirTran Airways	\N	FL	TRS	CITRUS	United States	Y
1442	Bemidji Airlines	\N	CH	BMJ	BEMIDJI	United States	Y

only showing top 20 rows

## Country or territory with highest number of airports

```
In [11]: airports_dat.groupBy("Country").count().orderBy("count", ascending=False).show()
```

Country	count
United States	1512
Canada	430
Australia	334
Brazil	264
Russia	264
Germany	249
China	241
France	217
United Kingdom	167
India	148
Indonesia	145
Japan	123
South Africa	99
Argentina	96
Mexico	84
Italy	83
Iran	82
Sweden	77
Turkey	76
Colombia	75

only showing top 20 rows

## Top K cities with most incoming airlines

```
In [12]: # Count number of incoming flights based on routes
K = 5
```

```

Incoming_Flight_Count = routes_dat.groupby("Destination_Airport_ID").count()
Top_K_Destination = Incoming_Flight_Count.orderBy("count", ascending=False).limit(K)

# Create list of top K airports_IDs
Top_K_ID_List = [row[0] for row in Top_K_Destination.select("Destination_Airport_ID").collect()]

# Filter top K airports from airports_dat
Top_K_Airports = airports_dat.filter(airports_dat["Airport_ID"].isin(Top_K_ID_List))

# Sort top K airports by city
Top_K_Cities = [row[0] for row in Top_K_Airports.select("City").collect()]
print("Top " + str(K) + " cities with most incoming airlines:")
i = 1
for city in Top_K_Cities:
    print(i, end=". ")
    print(city)
    i += 1

```

Top 5 cities with most incoming airlines:

1. London
2. Paris
3. Beijing
4. Atlanta
5. Chicago

## Geospatial Analytics

### Closest airport to a city X's geospatial coordinate

In [ ]:

### Airport in each US states geospatial boundary

In [33]:

```

states = state_boundary_dat.selectExpr("s_name", "ST_GeomFromWKT(s_bound) as s_bound")
states.createOrReplaceTempView("states")

US_Airports_LongLat = United_States_Airports.select("Name", concat(col("Longitude"),
                                                                    lit("_"), col("Latitude")) as LongLat)

airports_geo = US_Airports_LongLat.selectExpr("Name", "ST_PointFromText(LongLat, '\_\'') as geo")
airports_geo.createOrReplaceTempView("airports_geo")

airports_in_state = spark.sql("select * from states s, airports_geo a where ST_Contains(s.s_bound, a.geo)")
airports_in_state.show()

```

21/11/14 22:58:18 WARN JoinQuery: UseIndex is true, but no index exists. Will build index on the fly.

s_name	s_bound	Name	LongLat
Alaska	POLYGON ((-141.02... Barter Island LRR...	POINT (-143.58200...	
Alaska	POLYGON ((-141.02... Wainwright Air St...	POINT (-159.86000...	
Alaska	POLYGON ((-141.02... Cape Lisburne LRR...	POINT (-166.11000...	
Alaska	POLYGON ((-141.02... Point Lay LRRS Ai...	POINT (-163.00500...	
Hawaii	POLYGON ((-154.62... Hilo Internationa...	POINT (-155.04800...	
Florida	POLYGON ((-87.605... Orlando Executive...	POINT (-81.332901...	
Alaska	POLYGON ((-141.02... Bettles Airport	POINT (-151.52900...	
Alaska	POLYGON ((-141.02... Clear Airport	POINT (-149.11999...	
Alaska	POLYGON ((-141.02... Indian Mountain L...	POINT (-153.70399...	

```
|      Alaska|POLYGON ((-141.02...|Fort Yukon Airport|POINT (-145.25 66...|
|      Alaska|POLYGON ((-141.02...|Sparrevohn LRRS A...|POINT (-155.57400...|
|      Alaska|POLYGON ((-141.02...|Bryant Army Heliport|POINT (-149.65299...|
|      Alaska|POLYGON ((-141.02...|Tatalina LRRS Air...|POINT (-155.97700...|
|      Alaska|POLYGON ((-141.02...|Cape Romanzof LRR...|POINT (-166.03900...|
|Massachusetts|POLYGON ((-72.778...|Laurence G Hansco...|POINT (-71.289001...|
|      Alaska|POLYGON ((-141.02...|Cape Newenham LRR...|POINT (-162.06300...|
|      Alaska|POLYGON ((-141.02...|Iliamna Airport|POINT (-154.91099...|
|      Alaska|POLYGON ((-141.02...|Platinum Airport|POINT (-161.82000...|
|      Alaska|POLYGON ((-141.02...|Big Mountain Airport|POINT (-155.25900...|
|      Michigan|POLYGON ((-88.371...|Oscoda Wurtsmith ...|POINT (-83.394096...|
+-----+-----+-----+-----+
```

only showing top 20 rows

In [ ]: