SQL Analysis Documentation

Data Preparation and Processing

Preview the dataset busiest airports after loading into BigQuery to see what we are working with

```
#preview the data
SELECT *
FROM `movement-patterns.airports.incoming_arrivals`
LIMIT 10
```

Check for missing/Null values in the table

```
#check for NUll values
SELECT *
FROM `movement-patterns.airports.bussiest_airports`
WHERE Airport IS NULL
AND Code IS NULL
AND Location IS NULL
AND Country IS NULL
AND Passengers IS NULL
AND Rank IS NULL
```

Preview the dataset incoming arrivals after loading into BigQuery to check what we are working with

```
#preview the data
SELECT *
FROM `movement-patterns.airports.incoming_arrivals`
Limit 10
```

Preview the dataset incoming arrivals after loading into BigQuery to check what we are working with

```
#preview the data
SELECT *
FROM `movement-patterns.airports.departures`
Limit 10
```

From the preview we can see there are some missing values for the years.We shall save a new table with the data that has no null values to work with.

```
#removing null values in the years columns
select *
from `movement-patterns.airports.incoming_arrivals`
where coalesce(year_1995, year_1996, year_1997, year_1998, year_1999, year_2000,
year_2001, year_2002, year_2003, year_2004, year_2005, year_2006, year_2007,
year_2008, year_2009, year_2010, year_2011, year_2012, year_2013, year_2014,
year_2015, year_2016, year_2017, year_2018, year_2019, year_2020) IS NOT NULL
```

We shall save a new table with the data that has no null values to work with from the departures table.

```
#removing null values in the years columns
select *
from `movement-patterns.airports.departures`
where coalesce(year_1995, year_1996, year_1997, year_1998, year_1999, year_2000,
year_2001, year_2002, year_2003, year_2004, year_2005, year_2006, year_2007,
year_2008, year_2009, year_2010, year_2011, year_2012, year_2013, year_2014,
year_2015, year_2016, year_2017, year_2018, year_2019, year_2020) IS NOT NULL
```

Data Analysis

1. Checking summary statistics in the dataset

```
SELECT
  SUM(Passengers) AS TOTAL ,
  AVG(Passengers) AS AVERAGE,
  MAX(Passengers) AS MAXIMUM,
  MIN(Passengers) AS MINIMUM
FROM `movement-patterns.airports.bussiest_airports`
```

TOTAL	AVERAGE	MAXIMUM	MINIMUM
4296138973	42961389.730000019	103902992	23129400

2. Which is the busiest airport in the world?

```
SELECT *
FROM `movement-patterns.airports.bussiest_airports`
ORDER BY Passengers DESC
LIMIT 5
```

Rank	Airport	Code	Location	Country	Passengers
1	Hartsfield–Jackson Atlanta International Airport	ATL	Atlanta	United States	103902992
2	Beijing Capital International Airport	PEK	Beijing	China	95786442
3	Dubai International Airport	DXB	Dubai	United Arab Emirates	88242099
4	Los Angeles International Airport	LAX	Los Angeles	United States	84557968

5	O'Hare International Airport	ORD	Chicago	United States	79828183

This means Hartsfield_jackson airport handles the largest amount of passengers.

3. Which is the busiest country based on the number of passengers it handles?

```
SELECT
  DISTINCT(Country) as country,
  SUM(Passengers) as total_passengers,
  COUNT(Airport) as No_airports
FROM `movement-patterns.airports.bussiest_airports`
GROUP BY country
ORDER BY total_passengers  DESC
LIMIT 5
```

country	total_passengers	No_airports
United States	1177426600	25
China	674504074	16
United Kingdom	177382573	4
Japan	168575215	4
India	135703034	3

The United States is the country with the highest number of passengers its handling and this could be due to the fact that it has a total of 25 airports that have a high number of passengers each airport handles.

4. Merging the data for busiest airports , incoming arrivals and outgoing departures to further analyze the trends of movement into and out of the most busiest countries/airports.

```
SELECT
  airports.Country AS country, airports.total_passengers, airports.no_airports,
arrival.year_1995 AS arr_1995,arrival.year_1996 AS arr_1996,arrival.year_1997 AS
arr_1997,arrival.year_1998 AS arr_1998,arrival.year_2000 AS arr_2000,arrival.year_2001
AS arr_2001,arrival.year_2002 AS arr_2002,arrival.year_2003 AS arr_2003,
arrival.year_2004 AS arr_2004,arrival.year_2005 AS arr_2005,arrival.year_2006 AS
arr_2006,arrival.year_2007 AS arr_2007,arrival.year_2008 AS arr_2008,
```

```
arrival.year_2009 AS arr_2009,arrival.year_2010 AS arr_2010,arrival.year_2011 AS
arr_2011,arrival.year_2012 AS arr_2012,arrival.year_2013 AS arr_2013,arrival.year_2014
AS arr_2014,arrival.year_2015 AS arr_2015,arrival.year_2016 AS arr_2016,
arrival.year_2017 AS arr_2017,arrival.year_2018 AS arr_2018,arrival.year_2019 AS
arr_2019,arrival.year_2020 AS arr_2020,departures.year_1995 AS dep_1995,
departures.year_1996 AS dep_1996,departures.year_1997 AS dep_1997,
departures.year_1998 AS dep_1998, departures.year_1999 AS
dep_1999,departures.year_2000 AS dep_2000,departures.year_2001 AS dep_2001,
departures.year_2002 AS dep_2002,departures.year_2003 AS dep_2003,
departures.year_2004 AS dep_2004,departures.year_2005 AS dep_2005,departures.year_2006
AS dep_2006,departures.year_2007 AS dep_2007,departures.year_2008 AS dep_2008,
departures.year_2009 AS dep_2009,departures.year_2010 AS dep_2010,departures.year_2011
AS dep_2011,departures.year_2012 AS dep_2012,departures.year_2013 AS dep_2013,
departures.year_2014 AS dep_2014,departures.year_2015 AS dep_2015,
departures.year_2016 AS dep_2016,departures.year_2017 AS dep_2017,departures.year_2018
AS dep_2018,departures.year_2019 AS dep_2019
FROM ((`movement-patterns.airports.airports_per_country` AS airports
LEFT JOIN `movement-patterns.airports.incoming_arrivals_v2` AS arrival ON
airports.Country = arrival.country_name)
LEFT JOIN  `movement-patterns.airports.departures_v2` AS departures ON
airports.Country = departures.country_name)
ORDER BY no_airports DESC
```

5. Extracting from incoming arrivals the number of incoming passengers by region and economies.

```
#table of incoming arrivals by region
SELECT *
FROM `movement-patterns.airports.incoming_arrivals_v2`
WHERE country_name = 'World' OR
  country_name = 'Europe & Central Asia' OR
  country_name = 'North America' OR
  country_name = 'Middle East & North Africa' OR
  country_name = 'Sub-Saharan Africa' OR
  country_name = 'Latin America & Caribbean' OR
  country_name = 'South Asia' OR
  country_name = 'East Asia & Pacific'

#incoming arrivals by level of economy
SELECT *
FROM `movement-patterns.airports.incoming_arrivals_v2`
WHERE country_name = 'Low income' OR
```

```
  country_name = 'Lower middle income' OR
  country_name = 'Upper middle income' OR
  country_name = 'High income'
```

6. Extracting from departures number of outgoing passengers by regions and economies.

```
#outgoing departures by region
SELECT *
FROM `movement-patterns.airports.departures_v2`
WHERE country_name = 'World' OR
  country_name = 'Europe & Central Asia' OR
  country_name = 'North America' OR
  country_name = 'Middle East & North Africa' OR
  country_name = 'Sub-Saharan Africa' OR
  country_name = 'Latin America & Caribbean' OR
  country_name = 'South Asia' OR
  country_name = 'East Asia & Pacific'

#no data on middle east and north africa and sub saharan africa from 1995-2020.
#no data for year 2020 for all regions
ALTER TABLE `movement-patterns.airports.regions_departures`
DROP COLUMN year_2020

#outgoing departures by levels of economy
SELECT *
FROM `movement-patterns.airports.departures_v2`
WHERE country_name = 'Low income' OR
  country_name = 'Lower middle income' OR
  country_name = 'Upper middle income' OR
  country_name = 'High income'

#no data from lower and low income economies. This could be because they do not record
departures data or they do not have a lot of departures from those regions.
```