DATA 604 FINAL PROJECT DOCUMENTATION

NYC 2013 Flights Analysis

TEAM 1

Harshini Akkapally-RK84133

Soumya Kasireddy-GH35538

Showri Niharika Yeruva-VI29257

INDEX

ABSTRACT

New York is the busiest city in North America, with passengers flying to numerous destinations practically all year. Though the city has three major airports, Kennedy International Airport (JFK), Newark Liberty International Airport (EWR) and LaGuardia Airport (LGA), it is hard to manage the flights from this city because of the huge crowd. LGA is much closer to the city and handles smaller aircraft on shorter routes. JFK and EWR can handle larger aircraft and are better set up for large numbers of passengers, international arrivals, and transfers. But still, millions of travelers are affected by flight delays most often in New York city causing them to miss important meetings or miss connections and spend additional time away from home. Flight data analysis can help passengers to plan their travel and save time by avoiding delays. The analysis can also help to determine the average number of flights travelling to a city or to find the carrier with the best and worst on time records.

INTRODUCTION TO DATABASE SYSTEM

In this project, we would like to propose a database system that will create an easier and faster analysis method of NYC Flights. This database system consists of four different tables focusing on the different types of flights that fly from NYC from different airports. From tables that reviews the flights characteristics of the NYC the system will allow us to quickly come up with analytics of the delays and on-time records of the flights. Our project is focused on flights data; thus, we performed visualization and examined it using tableau.

DATASET:

SOURCE:

https://raw.githubusercontent.com/mguner/teaching/main/datasets/nyc_flights/nyc_flights.csv

- The dataset has nyc_flights.csv file containing the NYC flights data in 2013.
- It has an approximate of 336,776 flights in total.
- It has the detailed information of flights origin, destination, carriers.

Key Columns in the Flights Data Set:

- year, month, day : Date of departure.
- dep_time , arr_time : Actual departure and arrival times (format HHMM or HMM)
- sched_dep_time , sched_arr_time : Scheduled departure and arrival times (format HHMM or HMM)
- dep_delay , arr_delay : Departure and arrival delays, in minutes. Negative times represent early departures/arrivals.
- carrier : Two letter carrier abbreviation.
- Flight : Flight number
- Tailnum : Identification number of an aircraft painted on its tail
- Origin , dest : Origin and destination.
- air_time : Amount of time spent in the air, in minutes.

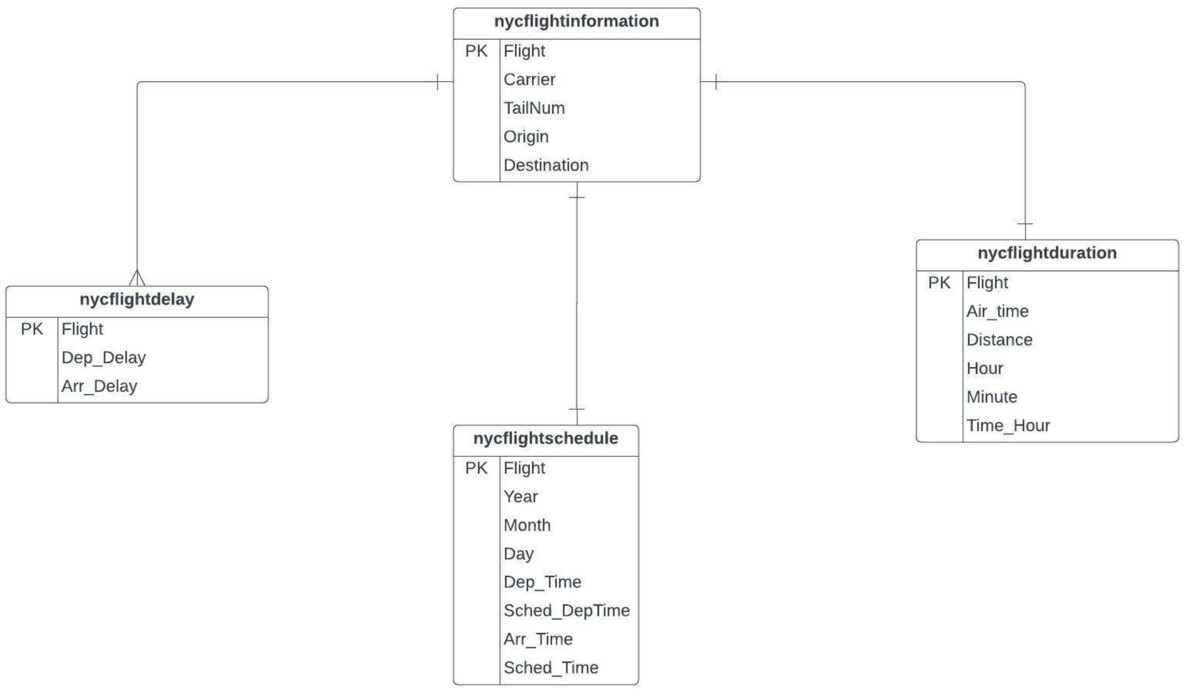
- Distance : Distance between airports, in miles.
- hour, minute : Time of scheduled departure broken into hour and minutes.
- time_hour : Scheduled date and hour of the flight as a POSIXct date

OBJECTIVES:

Our project involves the below objectives below:

- Flights that travel from certain origin / destination
- Flights that are operated by specific carriers
- Arrival or Departure delays of flights from different origin and destination
- Flights with shortest and longest duration
- Flights having arrival delay or departure delay more than certain time (2 hours)
- Which flights have worst on-time records from lowest to highest
- Flights availability in specific month to specific Destination

ER DIAGRAM:



PROCESS AND TOOLS USED:

AWS S3 Bucket: To load the flights_data.csv file loaded in the cloud storage and to store all the output files in the cloud

AWS RDS: To host the SQL Server

AWS Glue: To access the data and create table definitions

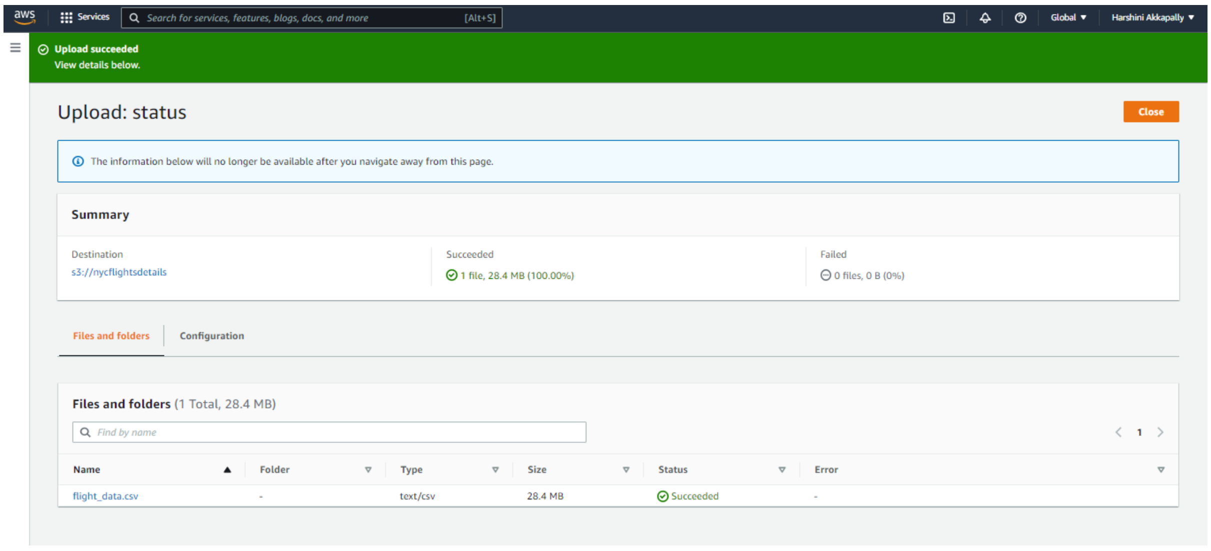
AWS Athena: To analyze data in Amazon S3 using standard SQL

Tableau: For Visualization and Analysis

AWS:

AWS S3 (Simple Storage Structure):

Created a S3 Bucket for uploading the flights_data.csv



Upload succeeded
View details below.

Upload: status

The information below will no longer be available after you navigate away from this page.

Summary

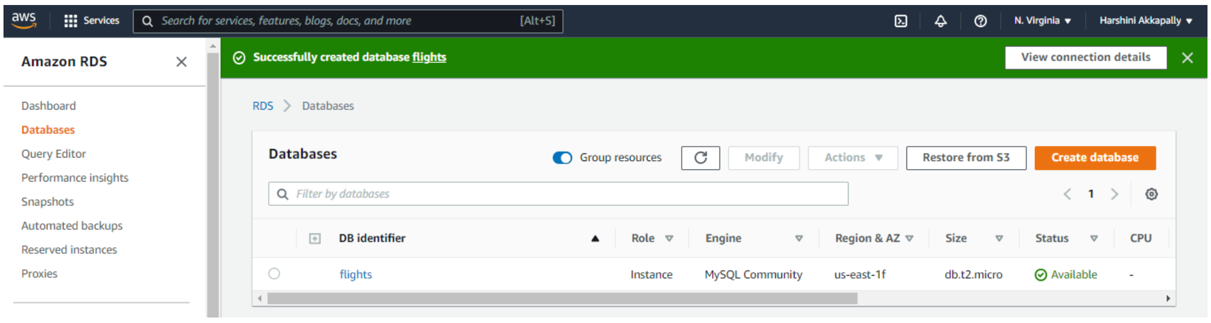
Destination	Succeeded	Failed
s3://nycflightsdetails	1 file, 28.4 MB (100.00%)	0 files, 0 B (0%)

Files and folders (1 Total, 28.4 MB)

Name	Folder	Type	Size	Status	Error
flight_data.csv	-	text/csv	28.4 MB	Succeeded	-

AWS RDS:

Created a database in AWS RDS



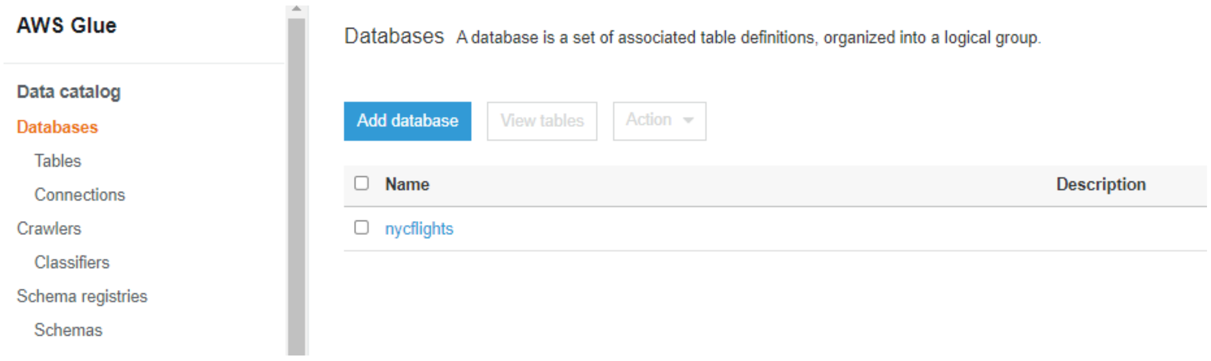
Successfully created database flights

Databases

DB identifier	Role	Engine	Region & AZ	Size	Status	CPU
flights	Instance	MySQL Community	us-east-1f	db.t2.micro	Available	-

AWS Glue:

Created a Database in AWS Glue

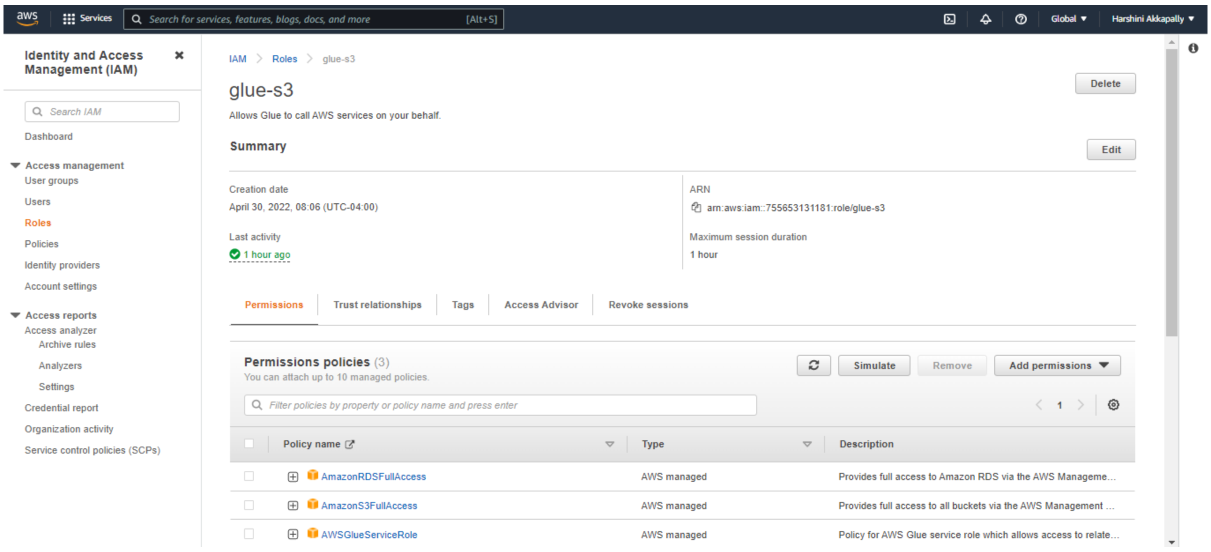


Databases A database is a set of associated table definitions, organized into a logical group.

Name	Description
nycflights	

Created a role ‘glue-s3’ in IAM (Identity and Access Management)

Provided Access to S3, RDS and AWS Glue



glue-s3

Allows Glue to call AWS services on your behalf.

Summary

Creation date
April 30, 2022, 08:06 (UTC-04:00)

Last activity
1 hour ago

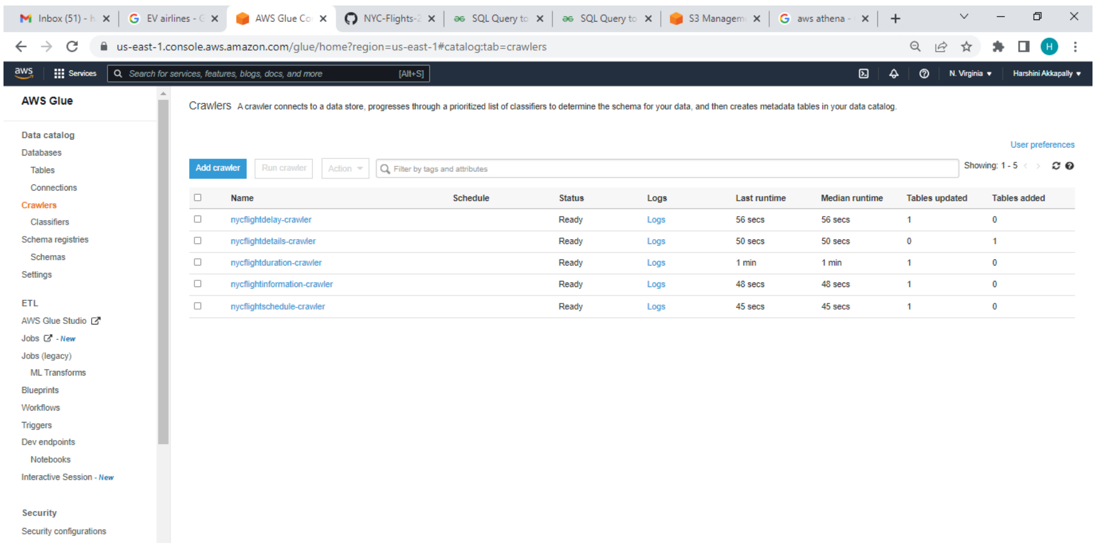
ARN
arn:aws:iam::755653131181:role/glue-s3

Maximum session duration
1 hour

Permissions policies (3)
You can attach up to 10 managed policies.

Policy name	Type	Description
AmazonRDSFullAccess	AWS managed	Provides full access to Amazon RDS via the AWS Manageme...
AmazonS3FullAccess	AWS managed	Provides full access to all buckets via the AWS Management ...
AWSGlueServiceRole	AWS managed	Policy for AWS Glue service role which allows access to relate...

Created Crawlers



us-east-1.console.aws.amazon.com/glue/home?region=us-east-1#catalog:tab=crawlers

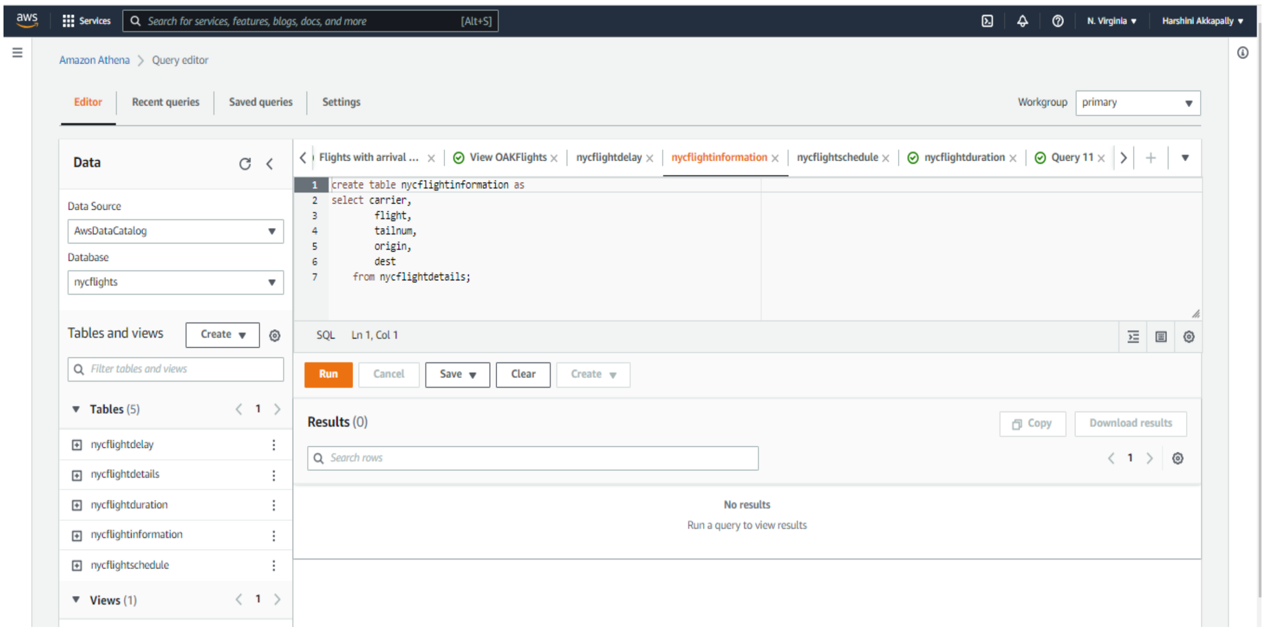
Crawlers A crawler connects to a data store, progresses through a prioritized list of classifiers to determine the schema for your data, and then creates metadata tables in your data catalog.

Name	Schedule	Status	Logs	Last runtime	Median runtime	Tables updated	Tables added
nycflightdelay-crawler		Ready	Logs	56 secs	56 secs	1	0
nycflightdetails-crawler		Ready	Logs	50 secs	50 secs	0	1
nycflightduration-crawler		Ready	Logs	1 min	1 min	1	0
nycflightinformation-crawler		Ready	Logs	48 secs	48 secs	1	0
nycflightschedule-crawler		Ready	Logs	45 secs	45 secs	1	0

AWS Athena:

Created Tables through Athena

create table nycflightinformation as select carrier, flight, tailnum, origin, dest from nycflightdetails;



Workgroup: primary

Data Source: AwsDataCatalog

Database: nycflights

```
create table nycflightinformation as
select carrier,
       flight,
       tailnum,
       origin,
       dest
    from nycflightdetails;
```

SQL Ln 1, Col 1

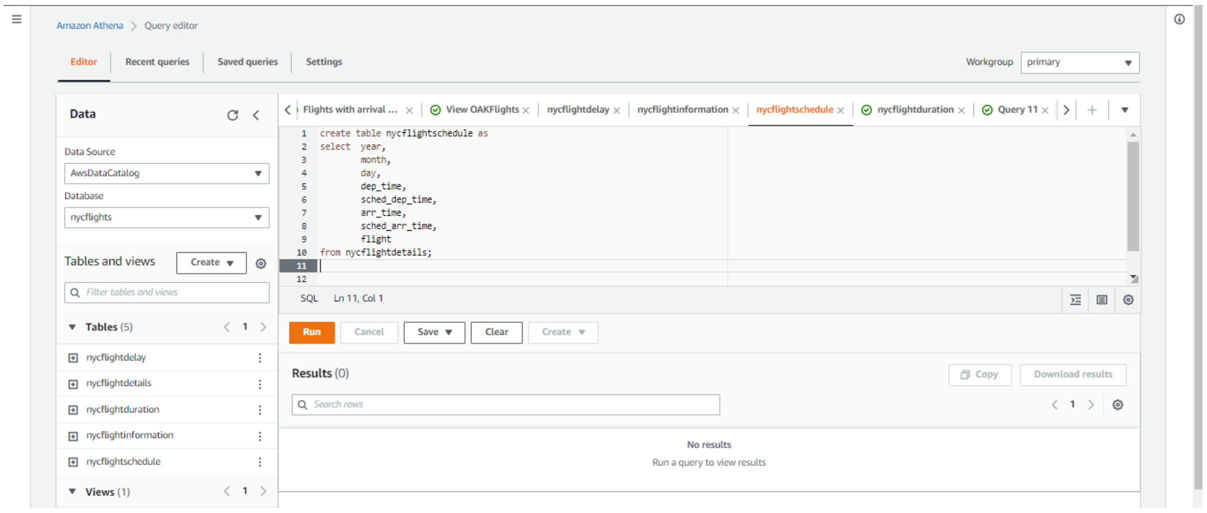
Tables (5)
- nycflightdelay
- nycflightdetails
- nycflightduration
- nycflightinformation
- nycflightschedule

Views (1)

Results (0)

No results
Run a query to view results

Create table nycflightschedule as select year, month, day, dep_time, sched_dep_time, arr_time, sched_arr_time, flight from nycflightdetails;



Workgroup: primary

Data Source: AwsDataCatalog

Database: nycflights

```
create table nycflightschedule as
select  year,
        month,
        day,
        dep_time,
        sched_dep_time,
        arr_time,
        sched_arr_time,
        flight
from nycflightdetails;
```

SQL Ln 11, Col 1

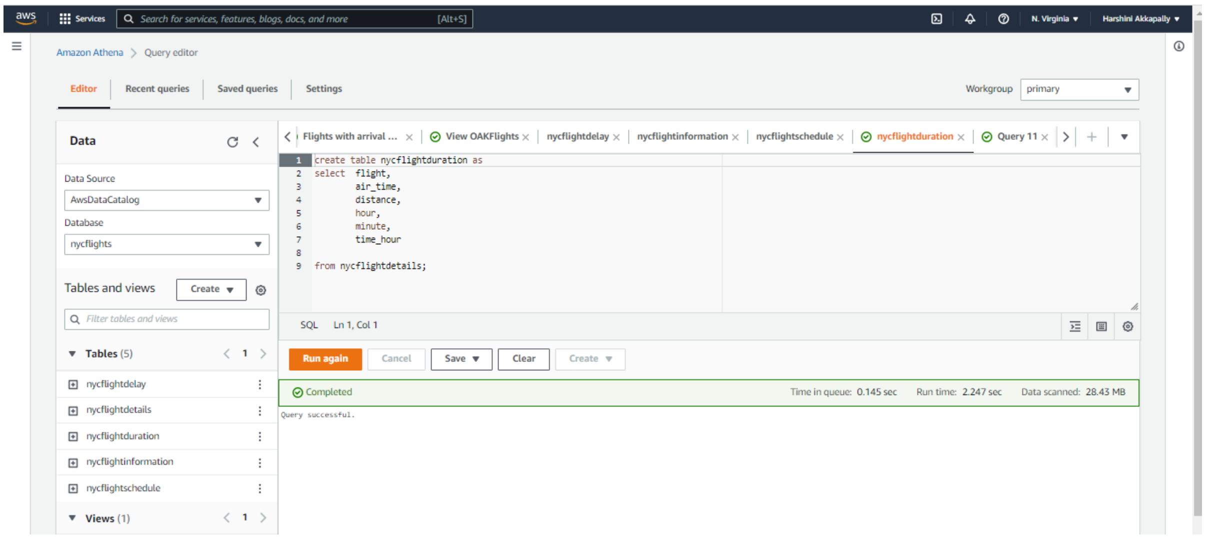
Tables (5)
- nycflightdelay
- nycflightdetails
- nycflightduration
- nycflightinformation
- nycflightschedule

Views (1)

Results (0)

No results
Run a query to view results

create table nycflightduration as select flight, air_time, distance, hour, minute, time_hour from nycflightdetails;



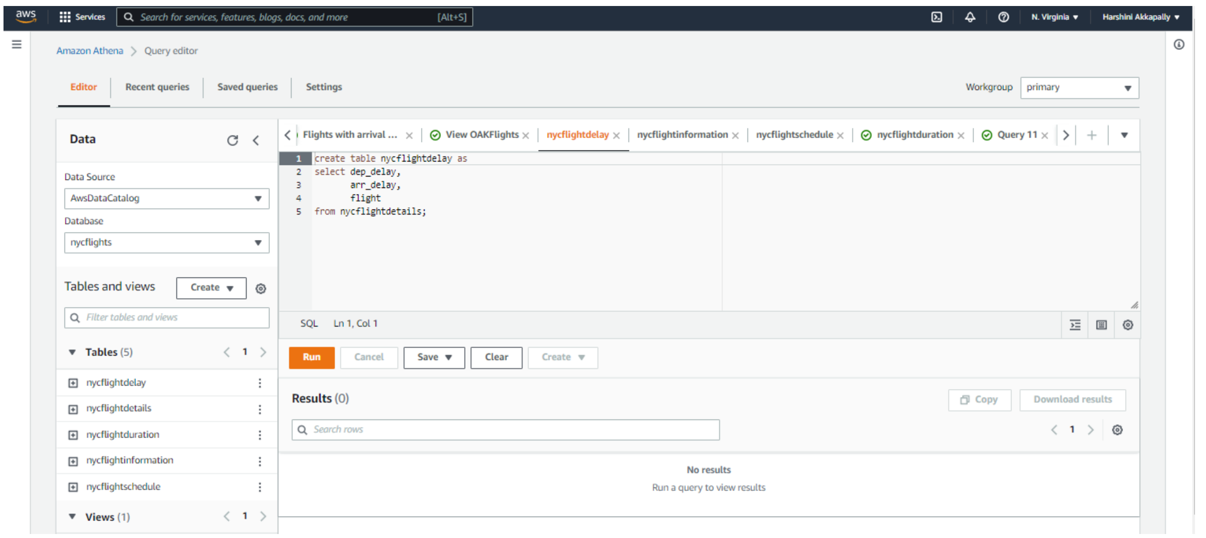
```
create table nycflightduration as
select  flight,
        air_time,
        distance,
        hour,
        minute,
        time_hour

from nycflightdetails;
```

Completed   Time in queue: 0.145 sec   Run time: 2.247 sec   Data scanned: 28.43 MB

```
Query successful.
```

create table nycflightdelay as select dep_delay, arr_delay, flight from nycflightdetails;

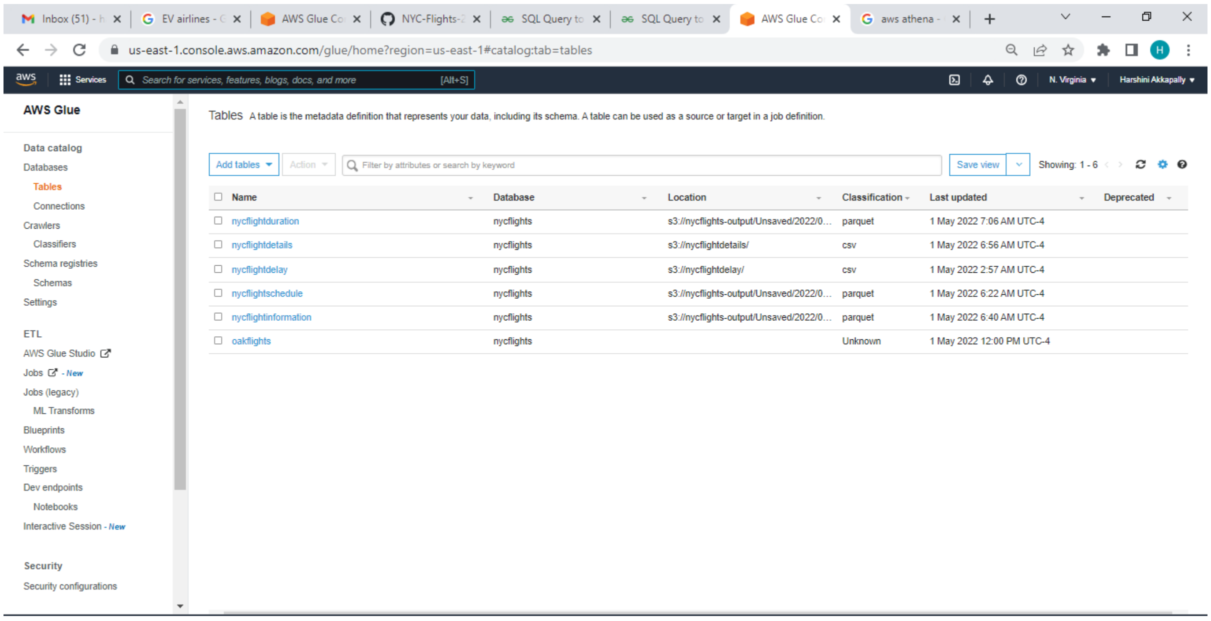


```
create table nycflightdelay as
select dep_delay,
       arr_delay,
       flight
from nycflightdetails;
```

Results (0)

No results
Run a query to view results

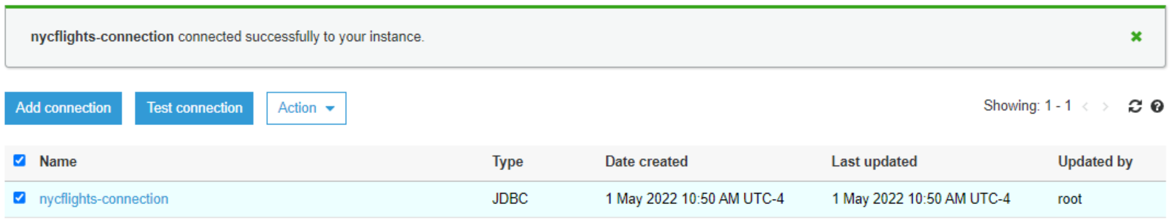
Tables which were Created after running Crawlers:



Tables A table is the metadata definition that represents your data, including its schema. A table can be used as a source or target in a job definition.

Name	Database	Location	Classification	Last updated	Deprecated
nycflightduration	nycflights	s3://nycflights-output/Unsaved/2022/0...	parquet	1 May 2022 7:06 AM UTC-4	
nycflightdetails	nycflights	s3://nycflightdetails/	csv	1 May 2022 6:56 AM UTC-4	
nycflightdelay	nycflights	s3://nycflightdelay/	csv	1 May 2022 2:57 AM UTC-4	
nycflightschedule	nycflights	s3://nycflights-output/Unsaved/2022/0...	parquet	1 May 2022 6:22 AM UTC-4	
nycflightinformation	nycflights	s3://nycflights-output/Unsaved/2022/0...	parquet	1 May 2022 6:40 AM UTC-4	
oakflights	nycflights		Unknown	1 May 2022 12:00 PM UTC-4	

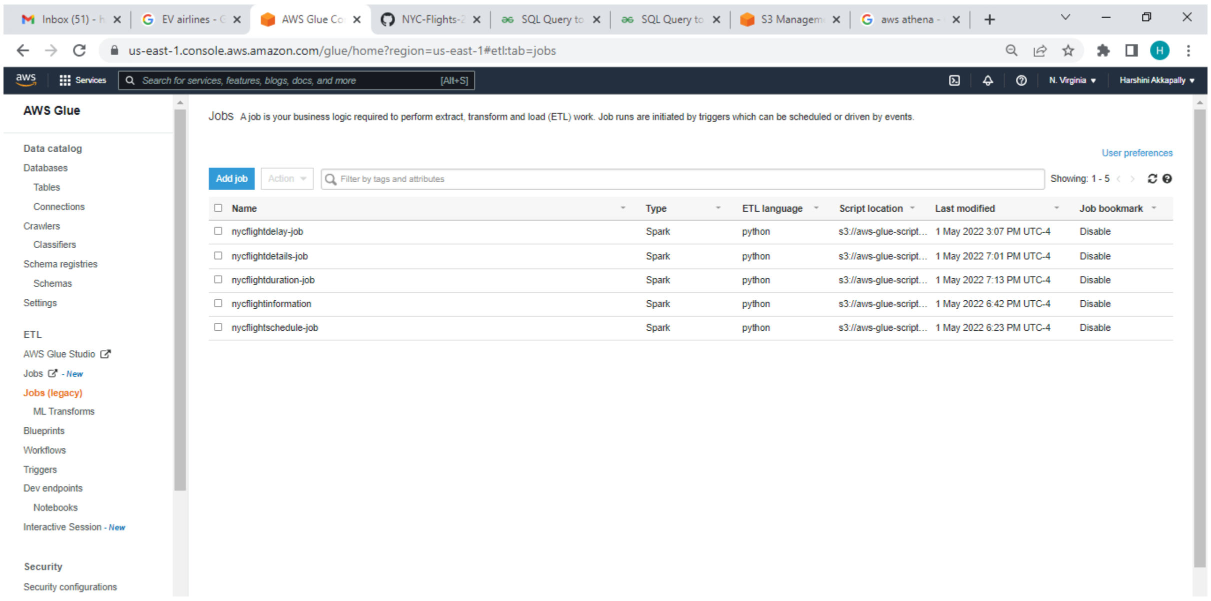
Created a Connection:



nycflights-connection connected successfully to your instance.

Name	Type	Date created	Last updated	Updated by
nycflights-connection	JDBC	1 May 2022 10:50 AM UTC-4	1 May 2022 10:50 AM UTC-4	root

Created Jobs in ETL:

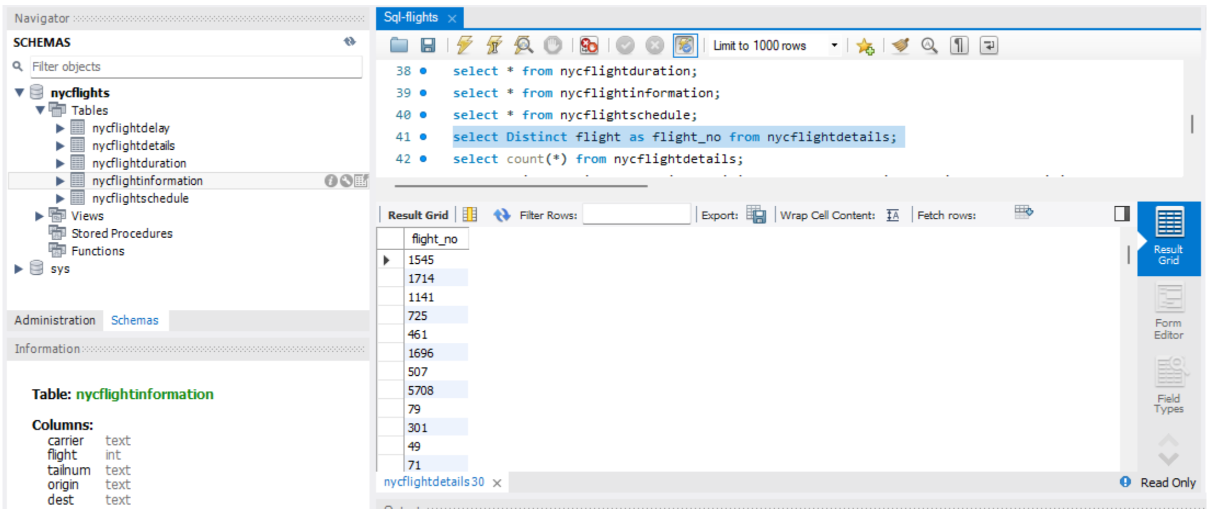


Jobs A job is your business logic required to perform extract, transform and load (ETL) work. Job runs are initiated by triggers which can be scheduled or driven by events.

Name	Type	ETL language	Script location	Last modified	Job bookmark
nycflightdelay-job	Spark	python	s3://aws-glue-script...	1 May 2022 3:07 PM UTC-4	Disable
nycflightdetails-job	Spark	python	s3://aws-glue-script...	1 May 2022 7:01 PM UTC-4	Disable
nycflightduration-job	Spark	python	s3://aws-glue-script...	1 May 2022 7:13 PM UTC-4	Disable
nycflightinformation	Spark	python	s3://aws-glue-script...	1 May 2022 6:42 PM UTC-4	Disable
nycflightschedule-job	Spark	python	s3://aws-glue-script...	1 May 2022 6:23 PM UTC-4	Disable

SQL QUERIES:

select Distinct flight as flight_no from nycflightdetails;



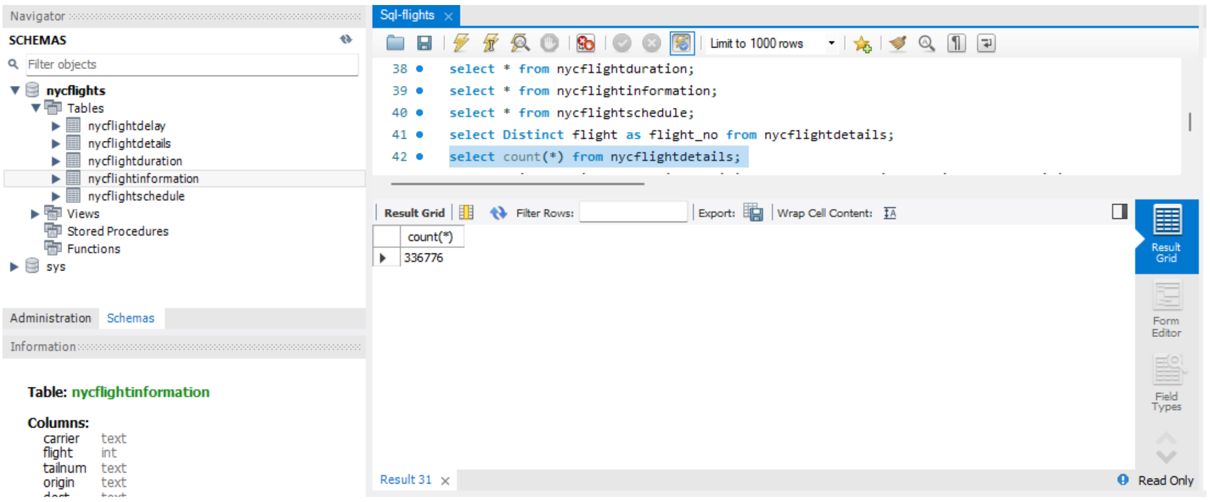
```
select * from nycflightduration;
select * from nycflightinformation;
select * from nycflightschedule;
select Distinct flight as flight_no from nycflightdetails;
select count(*) from nycflightdetails;
```

```
flight_no
1545
1714
1141
725
461
1696
507
5708
79
301
49
71
```

```
Table: nycflightinformation

Columns:
   carrier   text
   flight    int
   tailnum   text
   origin    text
   dest      text
```

select count(*) from nycflightdetails;



```
select * from nycflightduration;
select * from nycflightinformation;
select * from nycflightschedule;
select Distinct flight as flight_no from nycflightdetails;
select count(*) from nycflightdetails;
```

```
count(*)
336776
```

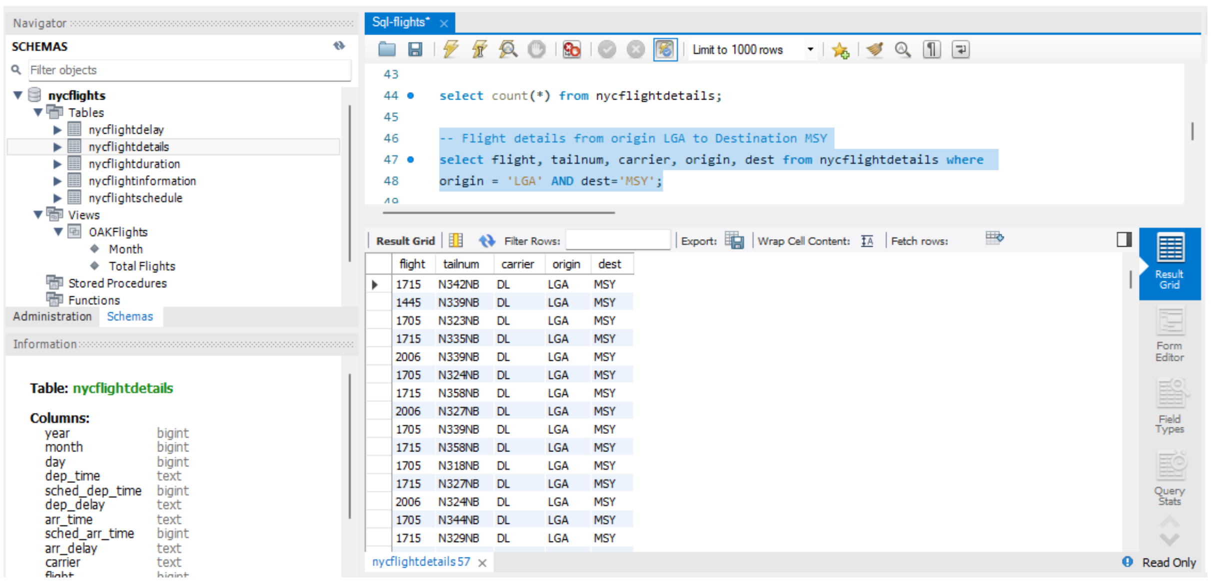
```
Table: nycflightinformation

Columns:
   carrier   text
   flight    int
   tailnum   text
   origin    text
```

-- Flight details from origin LGA to Destination MSY

select flight, tailnum, carrier, origin, dest from nycflightdetails where

origin = 'LGA' AND dest='MSY';



```
select count(*) from nycflightdetails;

-- Flight details from origin LGA to Destination MSY
select flight, tailnum, carrier, origin, dest from nycflightdetails where
origin = 'LGA' AND dest='MSY';
```

flight	tailnum	carrier	origin	dest
1715	N342NB	DL	LGA	MSY
1445	N339NB	DL	LGA	MSY
1705	N323NB	DL	LGA	MSY
1715	N335NB	DL	LGA	MSY
2006	N339NB	DL	LGA	MSY
1705	N324NB	DL	LGA	MSY
1715	N358NB	DL	LGA	MSY
2006	N327NB	DL	LGA	MSY
1705	N339NB	DL	LGA	MSY
1715	N358NB	DL	LGA	MSY
1705	N318NB	DL	LGA	MSY
1715	N327NB	DL	LGA	MSY
2006	N324NB	DL	LGA	MSY
1705	N344NB	DL	LGA	MSY
1715	N329NB	DL	LGA	MSY

Table: nycflightdetails

Columns:

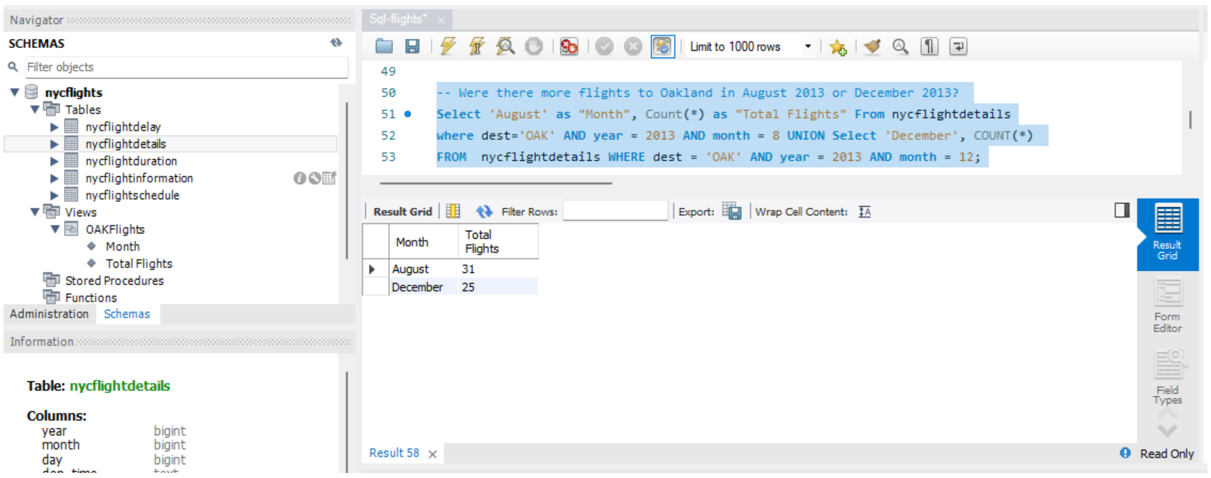
year	bigint
month	bigint
day	bigint
dep_time	text
sched_dep_time	bigint
dep_delay	text
arr_time	text
sched_arr_time	bigint
arr_delay	text
carrier	text

-- Were there more flights to Oakland in August 2013 or December 2013?

Select 'August' as "Month", Count(*) as "Total Flights" From nycflightdetails

where dest='OAK' AND year = 2013 AND month = 8 UNION Select 'December', COUNT(*)

FROM nycflightdetails WHERE dest = 'OAK' AND year = 2013 AND month = 12;



```
-- Were there more flights to Oakland in August 2013 or December 2013?
Select 'August' as "Month", Count(*) as "Total Flights" From nycflightdetails
where dest='OAK' AND year = 2013 AND month = 8 UNION Select 'December', COUNT(*)
FROM  nycflightdetails WHERE dest = 'OAK' AND year = 2013 AND month = 12;
```

Month	Total Flights
August	31
December	25

-- Creating View

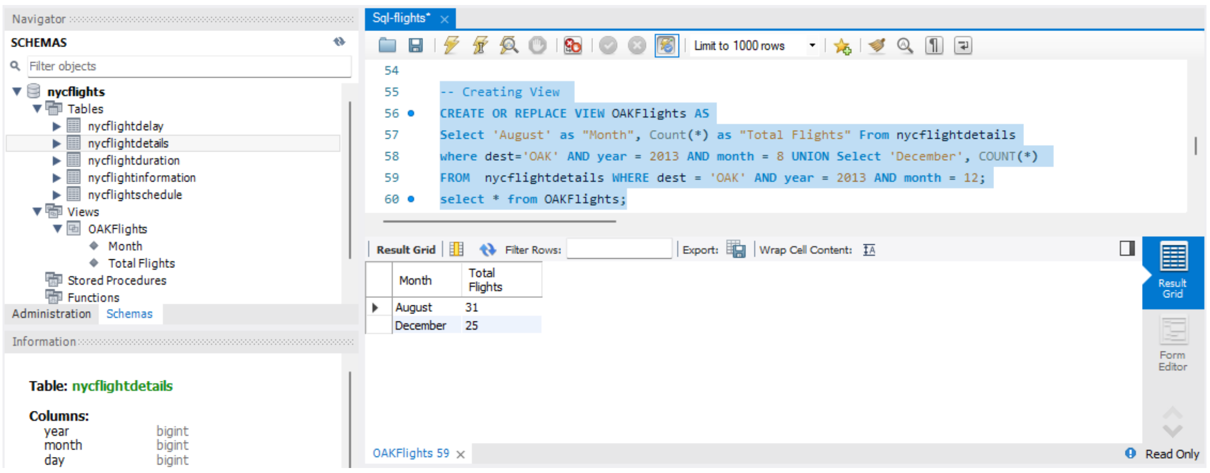
CREATE OR REPLACE VIEW OAKFlights AS

Select 'August' as "Month", Count(*) as "Total Flights" From nycflightdetails

where dest='OAK' AND year = 2013 AND month = 8 UNION Select 'December', COUNT(*)

FROM nycflightdetails WHERE dest = 'OAK' AND year = 2013 AND month = 12;

select * from OAKFlights;



```
-- Creating View
CREATE OR REPLACE VIEW OAKFlights AS
Select 'August' as "Month", Count(*) as "Total Flights" From nycflightdetails
where dest='OAK' AND year = 2013 AND month = 8 UNION Select 'December', COUNT(*)
FROM  nycflightdetails WHERE dest = 'OAK' AND year = 2013 AND month = 12;
select * from OAKFlights;
```

Month	Total Flights
August	31
December	25

Table: nycflightdetails

Columns:

year	bigint
month	bigint
day	bigint

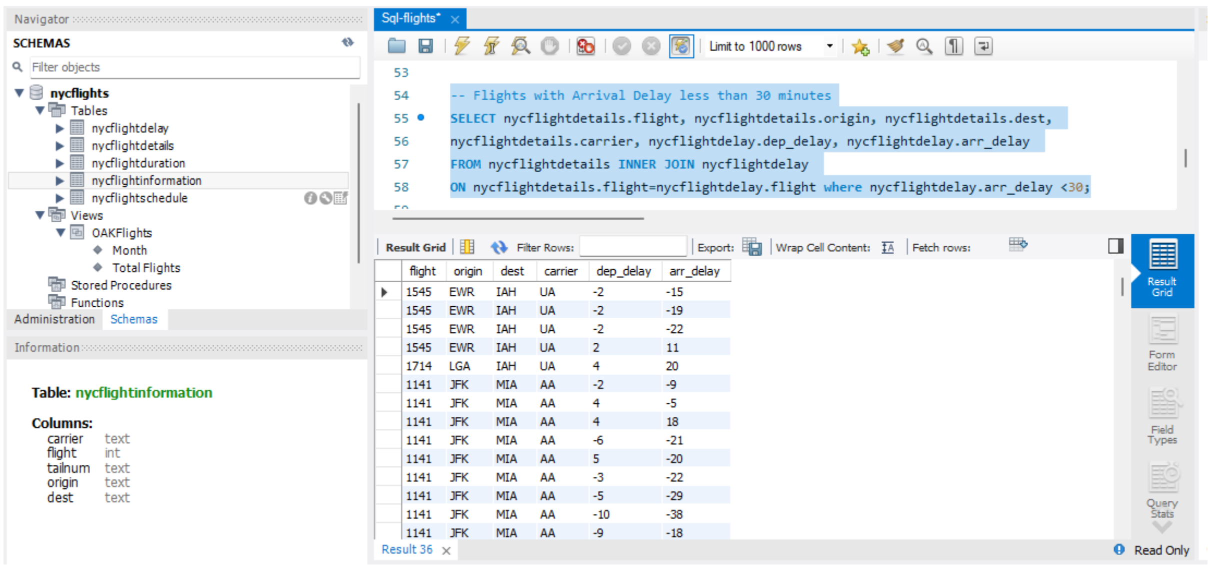
-- Flights with Arrival Delay less than 30 minutes

SELECT nycflightdetails.flight, nycflightdetails.origin, nycflightdetails.dest,

nycflightdetails.carrier, nycflightdelay.dep_delay, nycflightdelay.arr_delay

FROM nycflightdetails INNER JOIN nycflightdelay

ON nycflightdetails.flight=nycflightdelay.flight where nycflightdelay.arr_delay <30;



```sql
-- Flights with Arrival Delay less than 30 minutes
SELECT nycflightdetails.flight, nycflightdetails.origin, nycflightdetails.dest,
nycflightdetails.carrier, nycflightdelay.dep_delay, nycflightdelay.arr_delay
FROM nycflightdetails INNER JOIN nycflightdelay
ON nycflightdetails.flight=nycflightdelay.flight where nycflightdelay.arr_delay <30;
```

flight	origin	dest	carrier	dep_delay	arr_delay
1545	EWR	IAH	UA	-2	-15
1545	EWR	IAH	UA	-2	-19
1545	EWR	IAH	UA	-2	-22
1545	EWR	IAH	UA	2	11
1714	LGA	IAH	UA	4	20
1141	JFK	MIA	AA	-2	-9
1141	JFK	MIA	AA	4	-5
1141	JFK	MIA	AA	4	18
1141	JFK	MIA	AA	-6	-21
1141	JFK	MIA	AA	5	-20
1141	JFK	MIA	AA	-3	-22
1141	JFK	MIA	AA	-5	-29
1141	JFK	MIA	AA	-10	-38
1141	JFK	MIA	AA	-9	-18

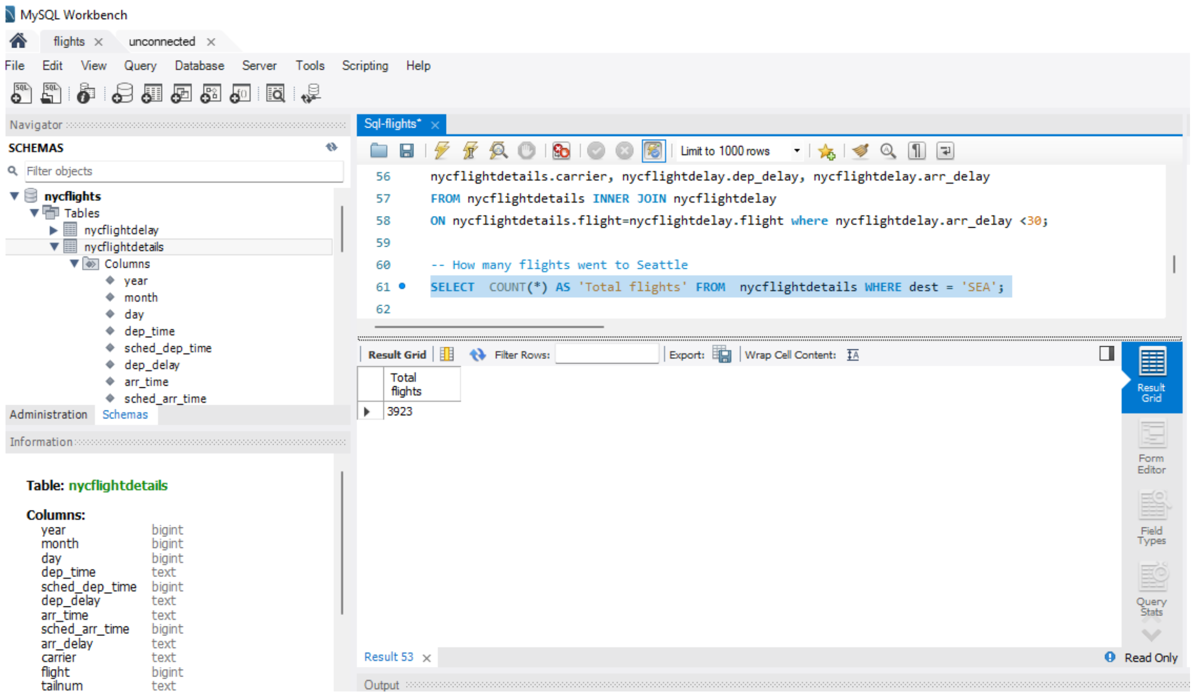
Table: nycflightinformation

Columns:

carrier	text
flight	int
tailnum	text
origin	text
dest	text

-- How many flights went to Seattle

SELECT COUNT(*) AS 'Total flights' FROM nycflightdetails WHERE dest = 'SEA';



```sql
nycflightdetails.carrier, nycflightdelay.dep_delay, nycflightdelay.arr_delay
FROM nycflightdetails INNER JOIN nycflightdelay
ON nycflightdetails.flight=nycflightdelay.flight where nycflightdelay.arr_delay <30;

-- How many flights went to Seattle
SELECT  COUNT(*) AS 'Total flights' FROM  nycflightdetails WHERE dest = 'SEA';
```

Total flights
3923

Table: nycflightdetails

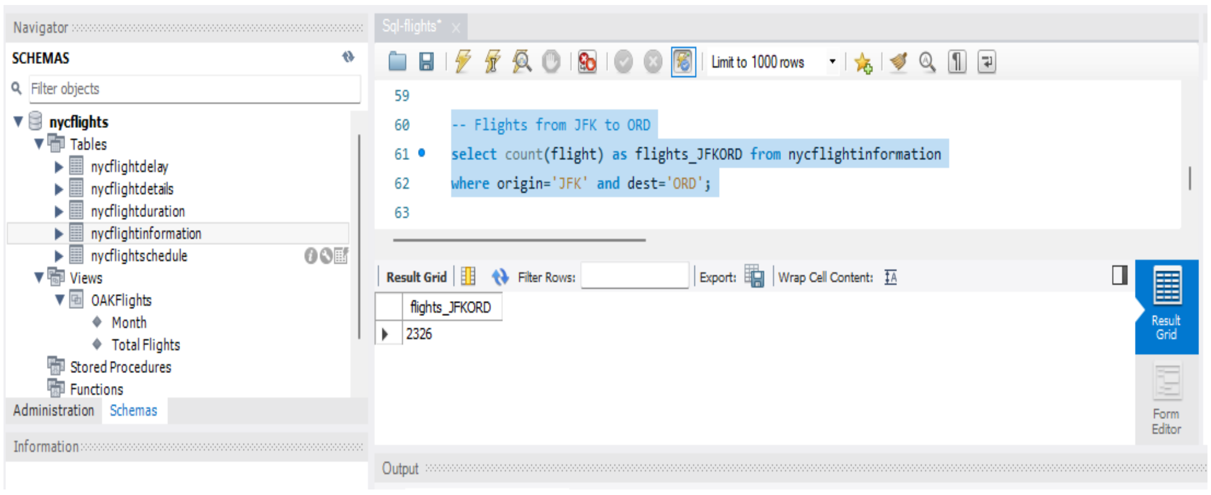
Columns:

year	bigint
month	bigint
day	bigint
dep_time	text
sched_dep_time	bigint
dep_delay	text
arr_time	text
sched_arr_time	bigint
arr_delay	text
carrier	text
flight	bigint
tailnum	text

-- Flights from JFK to ORD

select count(flight) as flights_JFKORD from nycflightinformation

where origin='JFK' and dest='ORD';



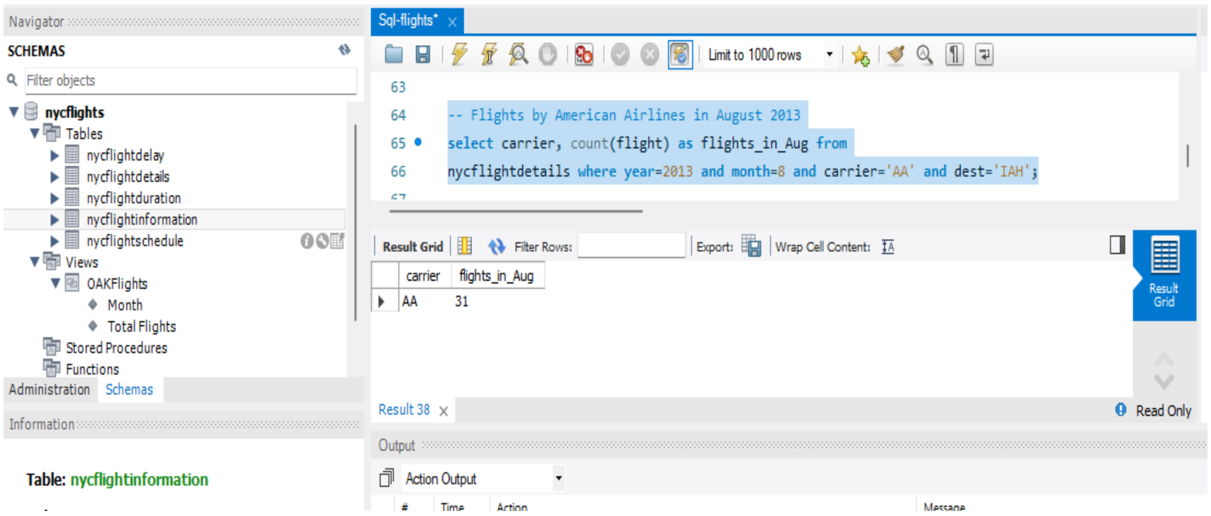
```
-- Flights from JFK to ORD
select count(flight) as flights_JFKORD from nycflightinformation
where origin='JFK' and dest='ORD';
```

```
flights_JFKORD
2326
```

-- Flights by American Airlines in August 2013

select carrier, count(flight) as flights_in_Aug from

nycflightdetails where year=2013 and month=8 and carrier='AA' and dest='IAH';

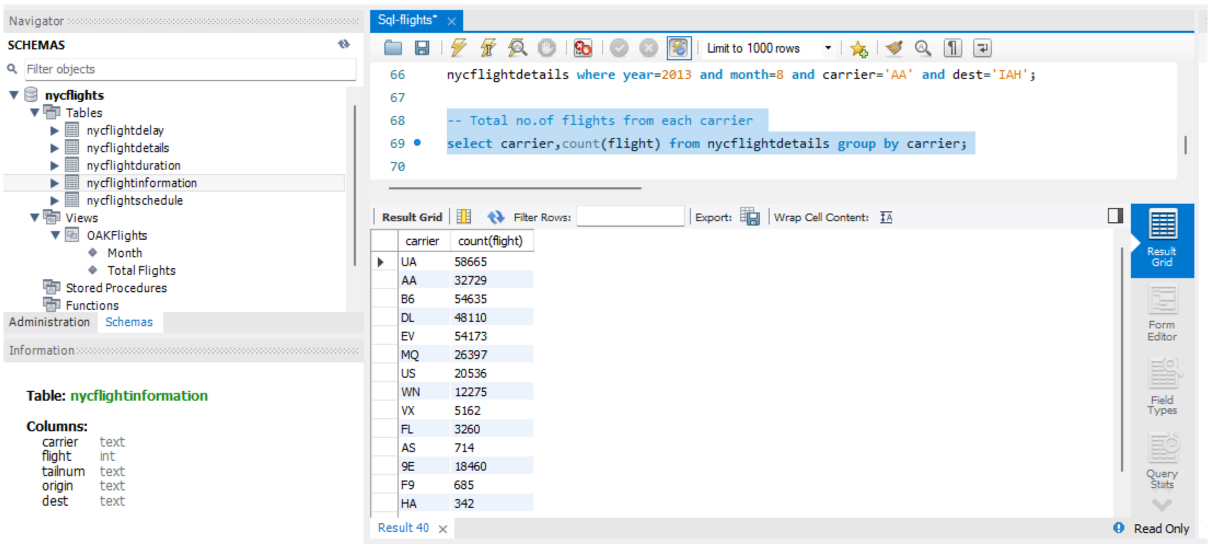


```
-- Flights by American Airlines in August 2013
select carrier, count(flight) as flights_in_Aug from
nycflightdetails where year=2013 and month=8 and carrier='AA' and dest='IAH';
```

```
carrier  flights_in_Aug
AA       31
```

-- Total no.of flights from each carrier

select carrier,count(flight) from nycflightdetails group by carrier;



```
nycflightdetails where year=2013 and month=8 and carrier='AA' and dest='IAH';

-- Total no.of flights from each carrier
select carrier,count(flight) from nycflightdetails group by carrier;
```

carrier	count(flight)
UA	58665
AA	32729
B6	54635
DL	48110
EV	54173
MQ	26397
US	20536
WN	12275
VX	5162
FL	3260
AS	714
9E	18460
F9	685
HA	342

Table: nycflightinformation

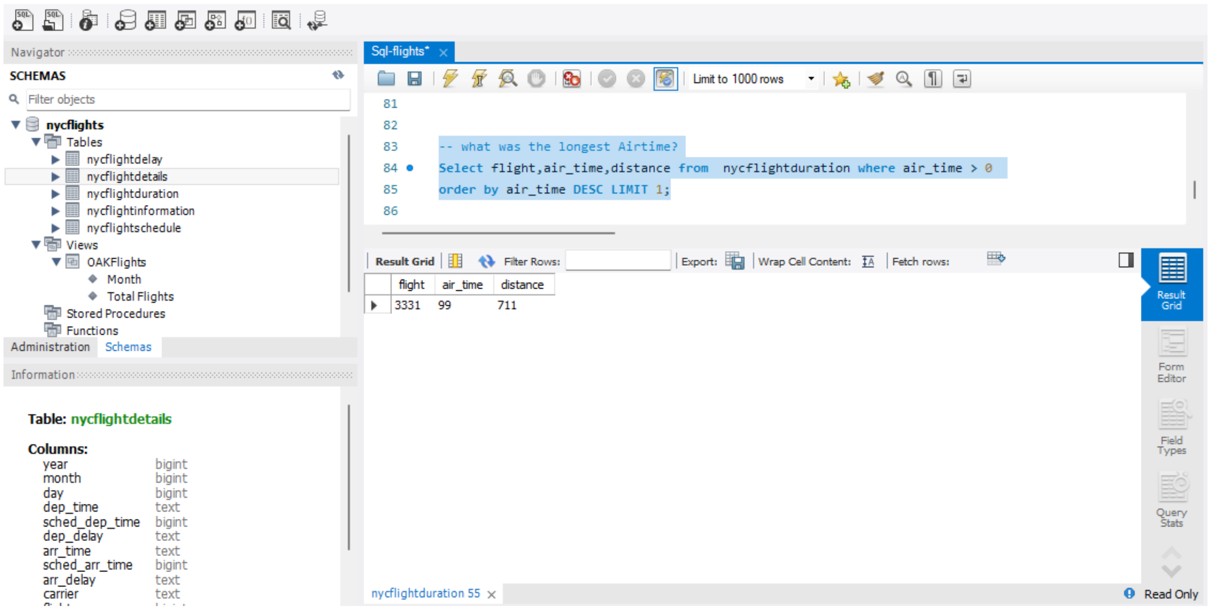
Columns:

carrier	text
flight	int
tailnum	text
origin	text
dest	text

-- what was the longest Airtime?

Select flight,air_time,distance from nycflightduration where air_time > 0

order by air_time DESC LIMIT 1;



```
-- what was the longest Airtime?
Select flight,air_time,distance from  nycflightduration where air_time > 0
order by air_time DESC LIMIT 1;
```

flight	air_time	distance
3331	99	711

Table: nycflightdetails

Columns:

year	bigint
month	bigint
day	bigint
dep_time	text
sched_dep_time	bigint
dep_delay	text
arr_time	text
sched_arr_time	bigint
arr_delay	text
carrier	text

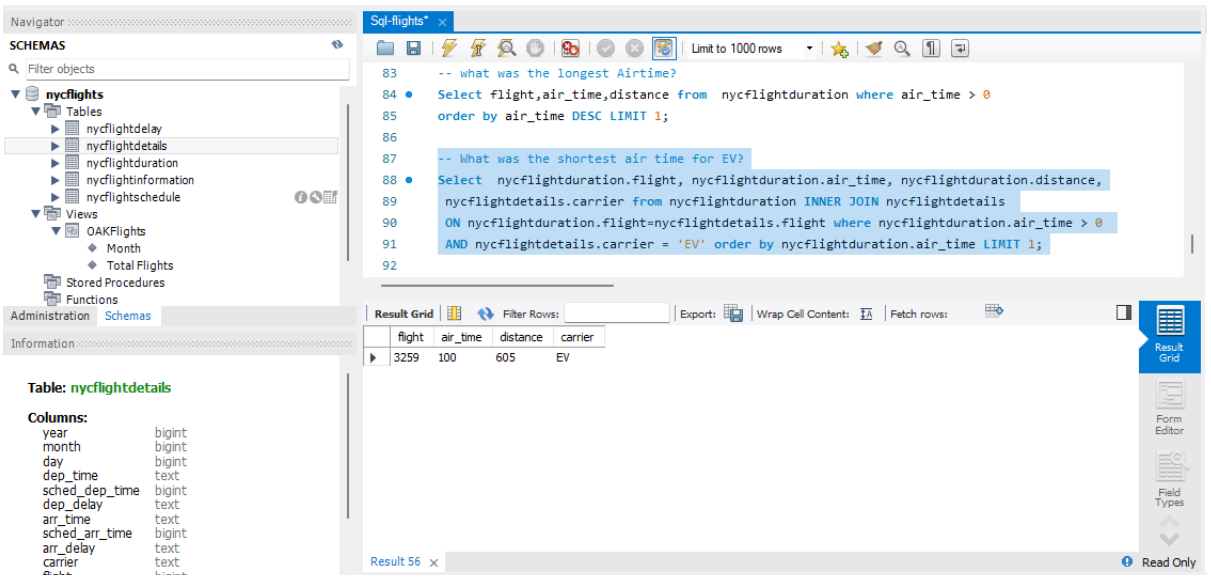
-- What was the shortest air time for EV?

Select nycflightduration.flight, nycflightduration.air_time, nycflightduration.distance,

 nycflightdetails.carrier from nycflightduration INNER JOIN nycflightdetails

 ON nycflightduration.flight=nycflightdetails.flight where nycflightduration.air_time > 0

 AND nycflightdetails.carrier = 'EV' order by nycflightduration.air_time LIMIT 1;



```
-- what was the longest Airtime?
Select flight,air_time,distance from  nycflightduration where air_time > 0
order by air_time DESC LIMIT 1;

-- What was the shortest air time for EV?
Select  nycflightduration.flight, nycflightduration.air_time, nycflightduration.distance,
 nycflightdetails.carrier from nycflightduration INNER JOIN nycflightdetails
 ON nycflightduration.flight=nycflightdetails.flight where nycflightduration.air_time > 0
 AND nycflightdetails.carrier = 'EV' order by nycflightduration.air_time LIMIT 1;
```

flight	air_time	distance	carrier
3259	100	605	EV

Table: nycflightdetails

Columns:

year	bigint
month	bigint
day	bigint
dep_time	text
sched_dep_time	bigint
dep_delay	text
arr_time	text
sched_arr_time	bigint
arr_delay	text
carrier	text

-- Arrival Delay and Depature Delay for flights from NC(EWR, LGA, JFK) to Seattle?

Select nycflightdelay.arr_delay, nycflightdelay.dep_delay, nycflightdetails.origin, nycflightdetails.dest AS 'Average Delay' from nycflightdelay INNER JOIN nycflightdetails

 ON nycflightdelay.flight=nycflightdetails.flight where nycflightdelay.arr_delay > 0

 AND nycflightdetails.dest='SEA' AND (nycflightdetails.origin='EWR' OR nycflightdetails.origin='JFK' OR nycflightdetails.origin='LGA');

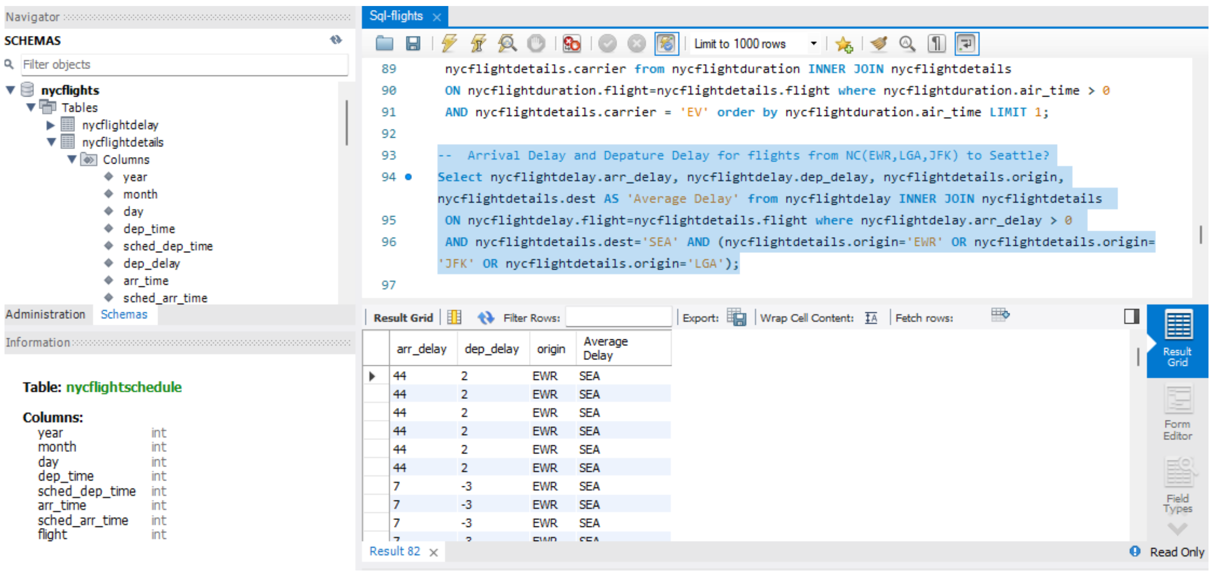


Table: nycflightschedule

Columns:

year	int
month	int
day	int
dep_time	int
sched_dep_time	int
arr_time	int
sched_arr_time	int
flight	int

```
     nycflightdetails.carrier from nycflightduration INNER JOIN nycflightdetails
     ON nycflightduration.flight=nycflightdetails.flight where nycflightduration.air_time > 0
     AND nycflightdetails.carrier = 'EV' order by nycflightduration.air_time LIMIT 1;

    --  Arrival Delay and Depature Delay for flights from NC(EWR,LGA,JFK) to Seattle?
    Select nycflightdelay.arr_delay, nycflightdelay.dep_delay, nycflightdetails.origin,
    nycflightdetails.dest AS 'Average Delay' from nycflightdelay INNER JOIN nycflightdetails
     ON nycflightdelay.flight=nycflightdetails.flight where nycflightdelay.arr_delay > 0
     AND nycflightdetails.dest='SEA' AND (nycflightdetails.origin='EWR' OR nycflightdetails.origin=
    'JFK' OR nycflightdetails.origin='LGA');
```

arr_delay	dep_delay	origin	Average Delay
44	2	EWR	SEA
44	2	EWR	SEA
44	2	EWR	SEA
44	2	EWR	SEA
44	2	EWR	SEA
44	2	EWR	SEA
7	-3	EWR	SEA
7	-3	EWR	SEA
7	-3	EWR	SEA
7	-3	EWR	SEA

-- What is the average arrival delay for flights from NC (EWR,LGA,JFK) to Seattle?

Select AVG(nycflightdelay.arr_delay) AS 'Average Arrival Delay' from nycflightdelay INNER JOIN nycflightdetails

ON nycflightdelay.flight=nycflightdetails.flight where nycflightdelay.arr_delay > 0

AND nycflightdetails.dest='SEA' AND (nycflightdetails.origin='EWR' OR nycflightdetails.origin='JFK' OR nycflightdetails.origin='LGA');

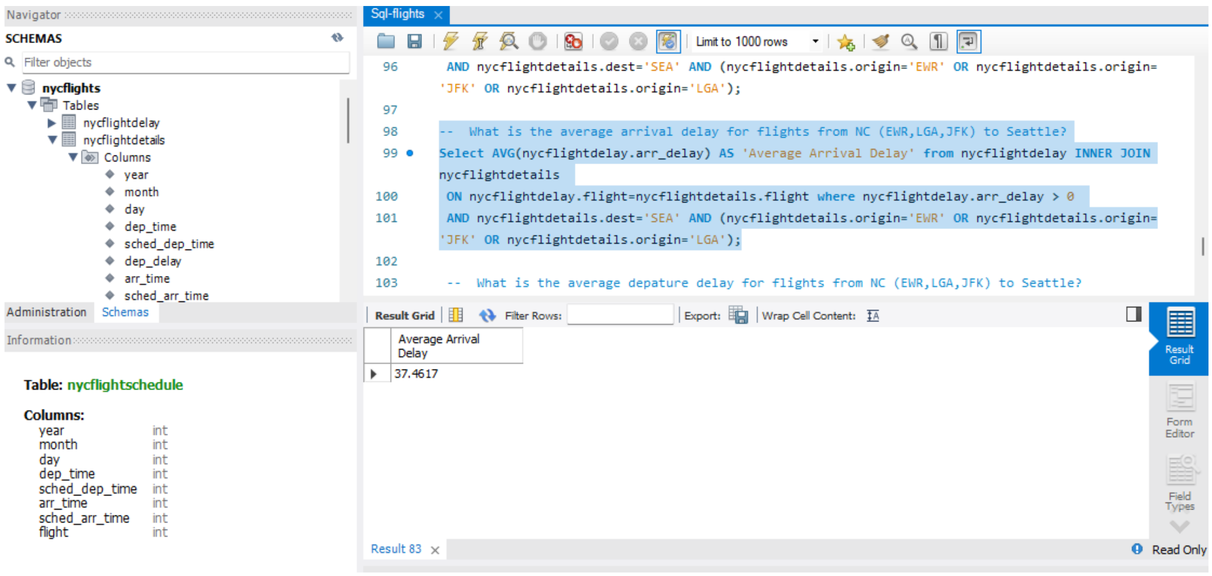


Table: nycflightschedule

Columns:

year	int
month	int
day	int
dep_time	int
sched_dep_time	int
arr_time	int
sched_arr_time	int
flight	int

```
     AND nycflightdetails.dest='SEA' AND (nycflightdetails.origin='EWR' OR nycflightdetails.origin=
    'JFK' OR nycflightdetails.origin='LGA');

    -- What is the average arrival delay for flights from NC (EWR,LGA,JFK) to Seattle?
    Select AVG(nycflightdelay.arr_delay) AS 'Average Arrival Delay' from nycflightdelay INNER JOIN
    nycflightdetails
     ON nycflightdelay.flight=nycflightdetails.flight where nycflightdelay.arr_delay > 0
     AND nycflightdetails.dest='SEA' AND (nycflightdetails.origin='EWR' OR nycflightdetails.origin=
    'JFK' OR nycflightdetails.origin='LGA');

     -- What is the average depature delay for flights from NC (EWR,LGA,JFK) to Seattle?
```

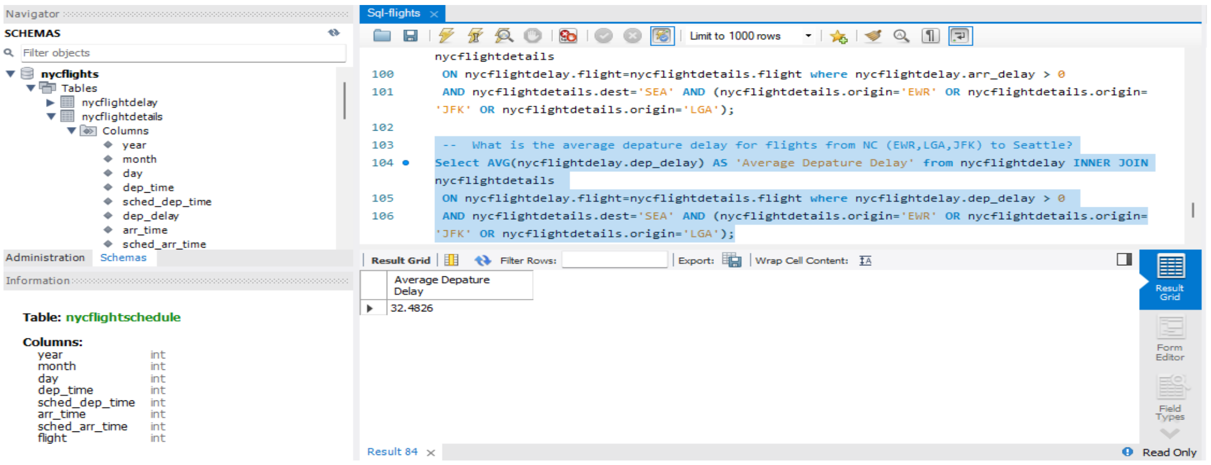
Average Arrival Delay
37.4617

-- What is the average depature delay for flights from NC (EWR,LGA,JFK) to Seattle?

Select AVG(nycflightdelay.dep_delay) AS 'Average Depature Delay' from nycflightdelay INNER JOIN nycflightdetails

ON nycflightdelay.flight=nycflightdetails.flight where nycflightdelay.dep_delay > 0

AND nycflightdetails.dest='SEA' AND (nycflightdetails.origin='EWR' OR nycflightdetails.origin='JFK' OR nycflightdetails.origin='LGA');



```
nycflightdetails
 ON nycflightdelay.flight=nycflightdetails.flight where nycflightdelay.arr_delay > 0
 AND nycflightdetails.dest='SEA' AND (nycflightdetails.origin='EWR' OR nycflightdetails.origin=
'JFK' OR nycflightdetails.origin='LGA');

 -- What is the average depature delay for flights from NC (EWR,LGA,JFK) to Seattle?
Select AVG(nycflightdelay.dep_delay) AS 'Average Depature Delay' from nycflightdelay INNER JOIN
nycflightdetails
 ON nycflightdelay.flight=nycflightdetails.flight where nycflightdelay.dep_delay > 0
 AND nycflightdetails.dest='SEA' AND (nycflightdetails.origin='EWR' OR nycflightdetails.origin=
'JFK' OR nycflightdetails.origin='LGA');
```

Average Depature Delay
32.4826

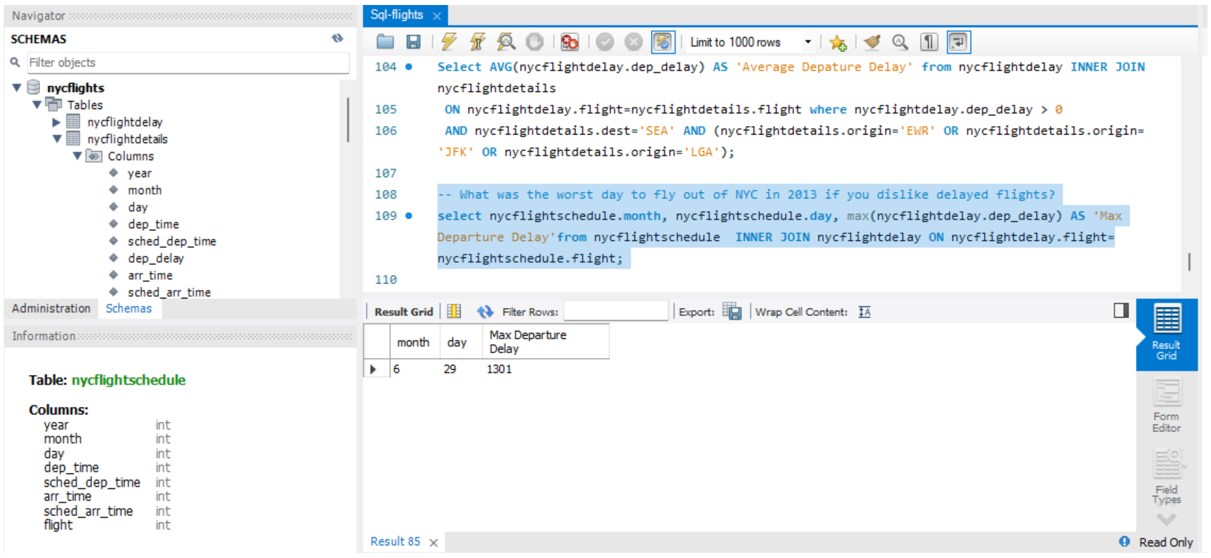
Table: nycflightschedule

Columns:

year	int
month	int
day	int
dep_time	int
sched_dep_time	int
arr_time	int
sched_arr_time	int
flight	int

-- What was the worst day to fly out of NYC in 2013 if you dislike delayed flights?

select nycflightschedule.month, nycflightschedule.day, max(nycflightdelay.dep_delay) AS 'Max Departure Delay'from nycflightschedule INNER JOIN nycflightdelay ON nycflightdelay.flight=nycflightschedule.flight;



```
Select AVG(nycflightdelay.dep_delay) AS 'Average Depature Delay' from nycflightdelay INNER JOIN
nycflightdetails
 ON nycflightdelay.flight=nycflightdetails.flight where nycflightdelay.dep_delay > 0
 AND nycflightdetails.dest='SEA' AND (nycflightdetails.origin='EWR' OR nycflightdetails.origin=
'JFK' OR nycflightdetails.origin='LGA');

-- What was the worst day to fly out of NYC in 2013 if you dislike delayed flights?
select nycflightschedule.month, nycflightschedule.day, max(nycflightdelay.dep_delay) AS 'Max
Departure Delay'from nycflightschedule  INNER JOIN nycflightdelay ON nycflightdelay.flight=
nycflightschedule.flight;
```

month	day	Max Departure Delay
6	29	1301

Table: nycflightschedule

Columns:

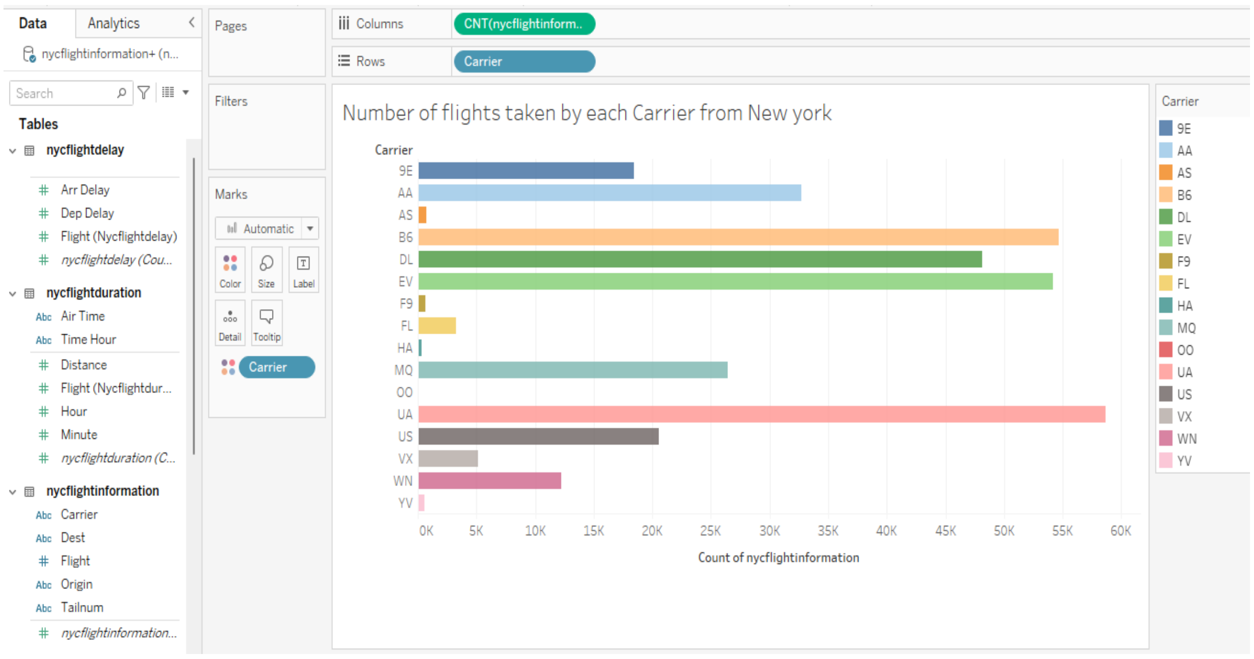
year	int
month	int
day	int
dep_time	int
sched_dep_time	int
arr_time	int
sched_arr_time	int
flight	int

DATA VISUALIZATION USING TABLEAU:

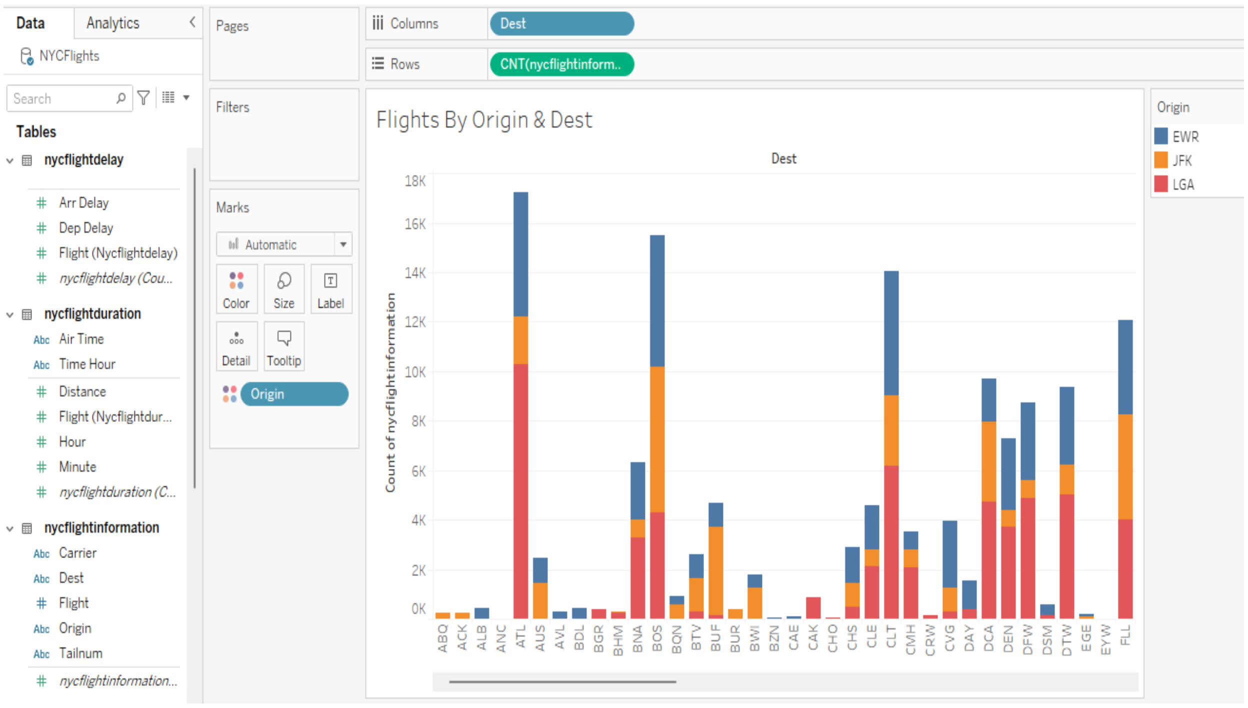
Data visualization is the act of converting an interpretation of data into a visual format, such as a guide or chart, to make information easier for the human mind to absorb and extract bits of knowledge from. The basic goal of data representation is to make it easier to recognize examples, patterns, and exceptions in large data sets. Data visualization is a sort of visual art that grabs our attention and keeps it there. We quickly see patterns and exceptions when we look at a diagram.

Tableau is a leading data visualization tool used for data analysis and business intelligence. It is a is a powerful tool for analysing large, complex datasets through a user-friendly interface.

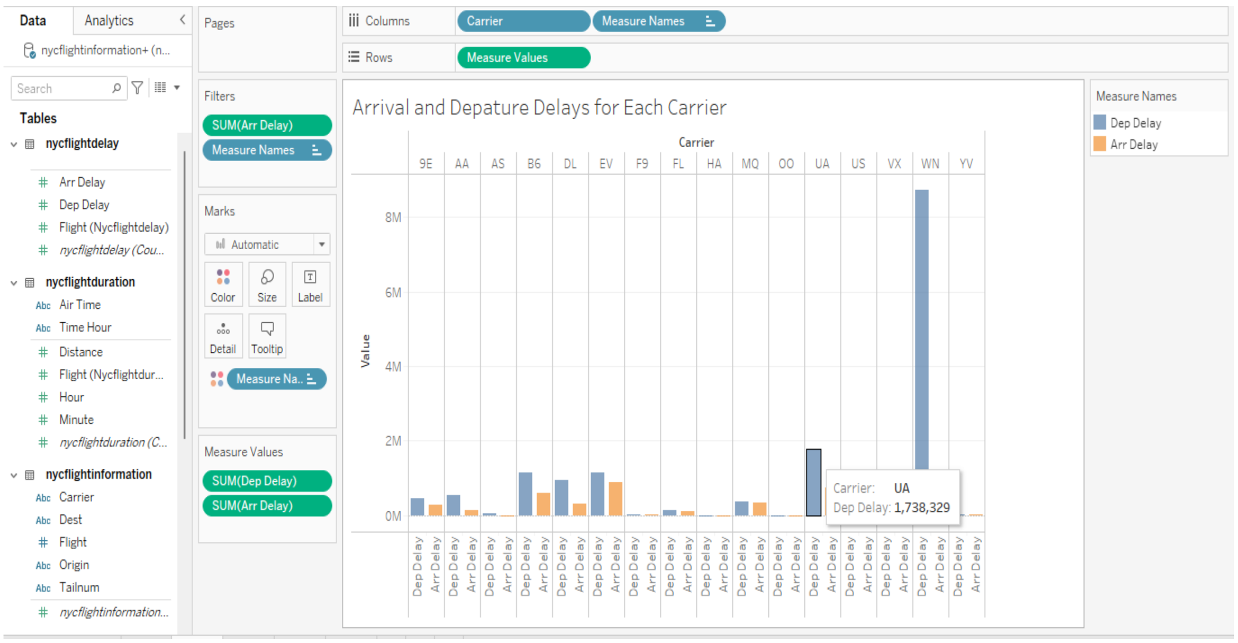
Number of flights taken by each Carrier from New York



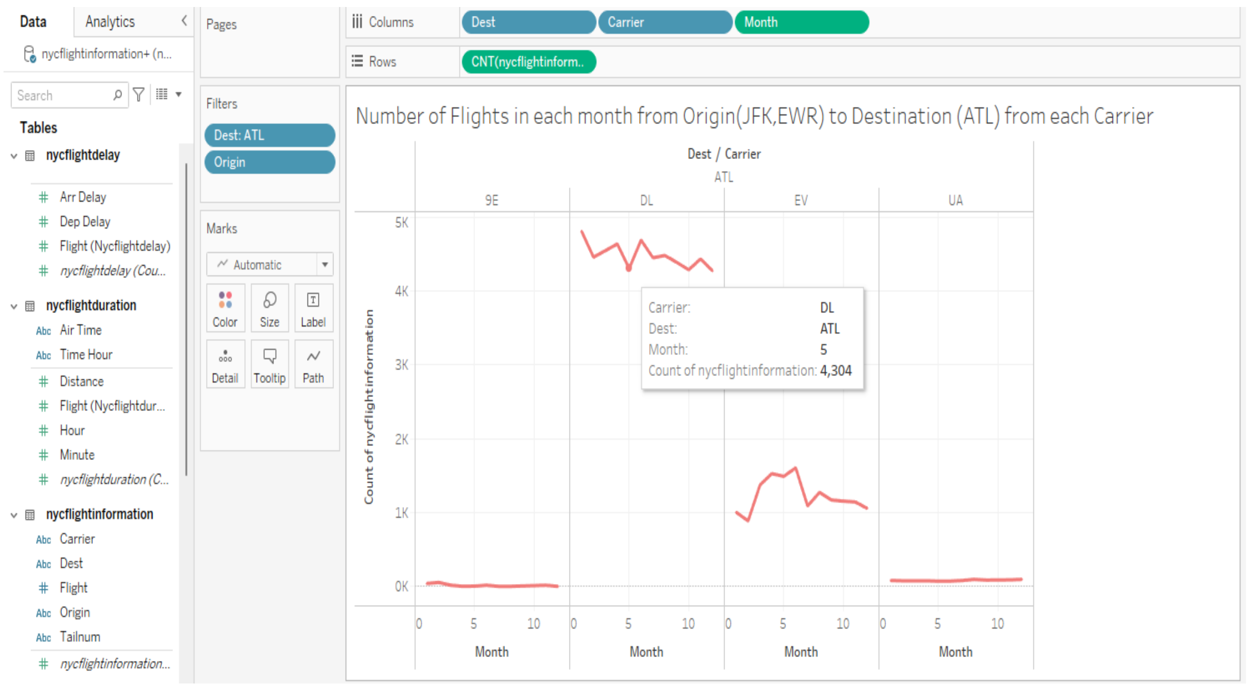
Flights By Origin & Destination



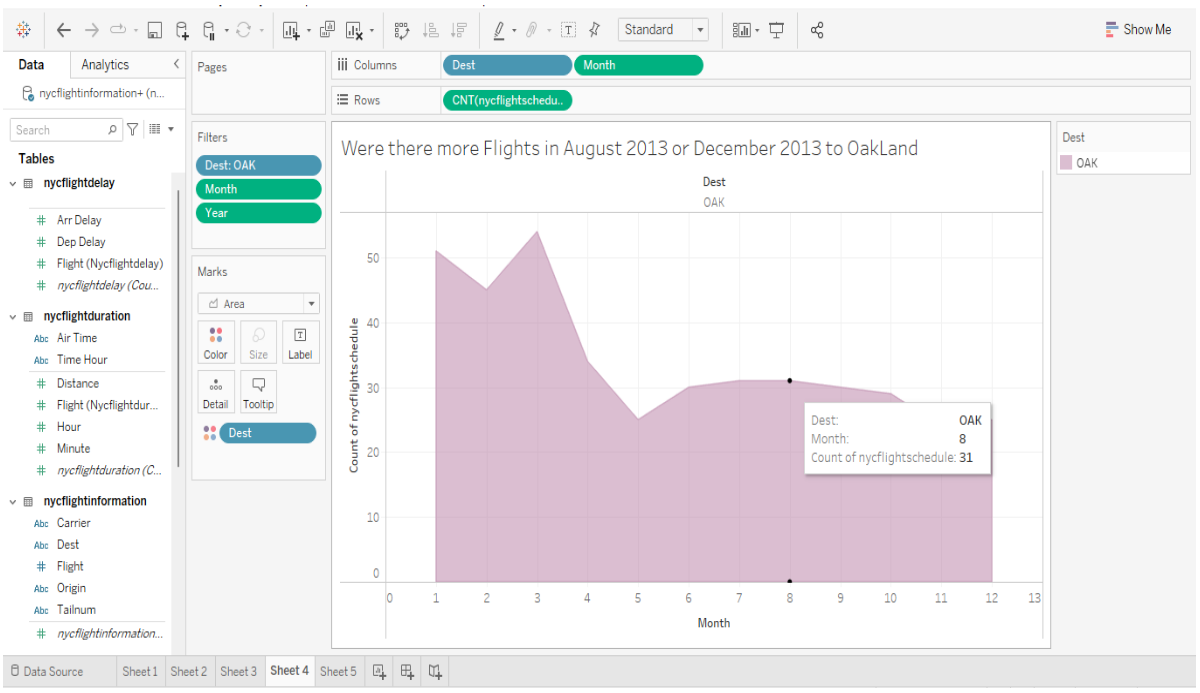
Arrival and Departure Delays for Each Carrier



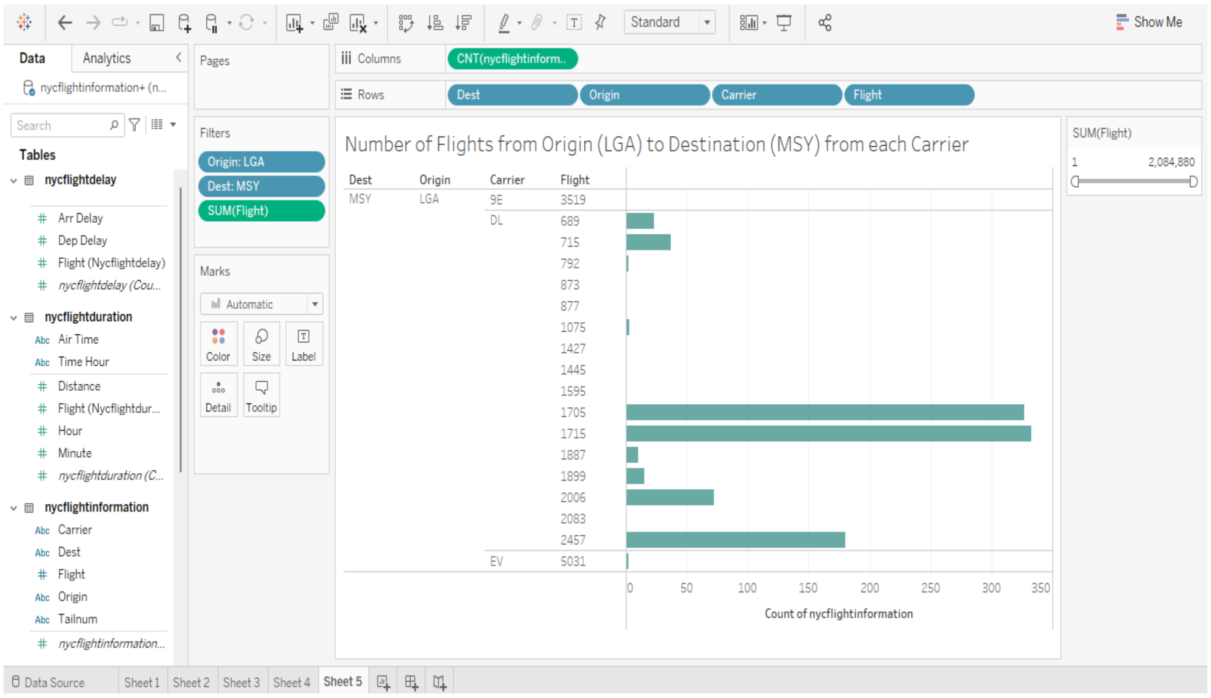
Number of Flights in each month from Origin(JFK,EWR) to Destination (ATL) from each Carrier



Were there more Flights in August 2013 or December 2013 to OakLand



Number of Flights from Origin (LGA) to Destination (MSY) from each Carrier



Columns: CNT(nycflightinform..

Rows: Dest | Origin | Carrier | Flight

Filters: Origin: LGA; Dest: MSY; SUM(Flight)

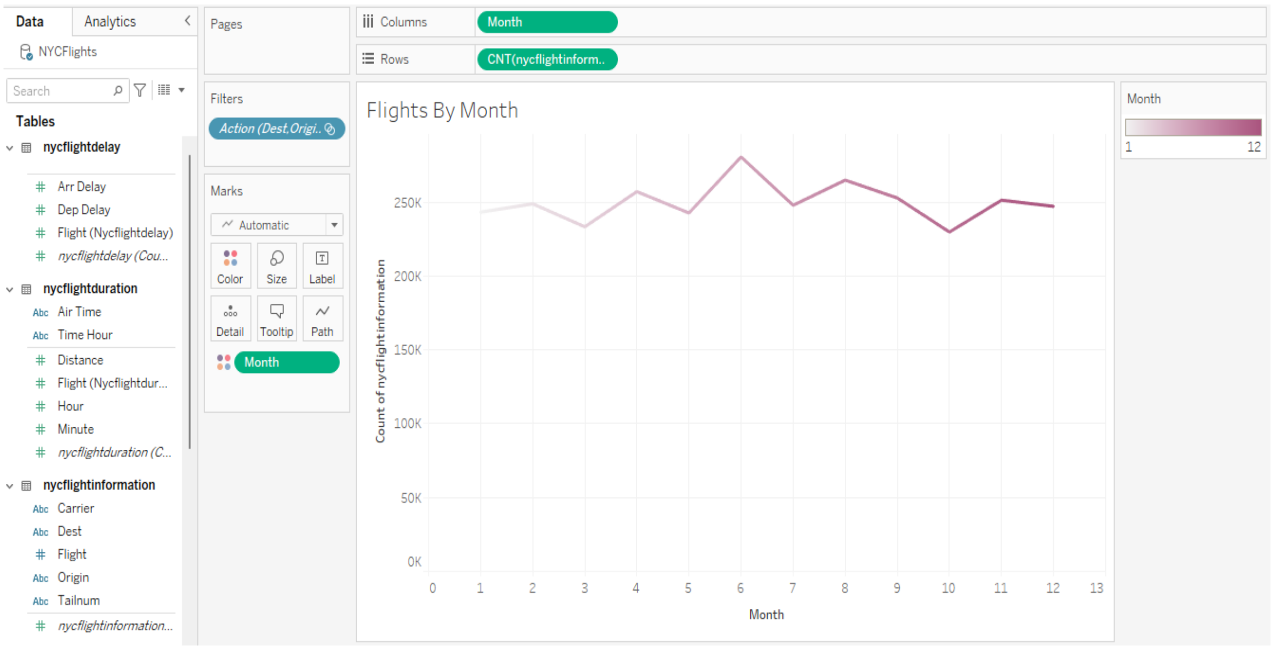
Marks: Automatic

SUM(Flight): 1 – 2,084,880

Number of Flights from Origin (LGA) to Destination (MSY) from each Carrier

Dest	Origin	Carrier	Flight
MSY	LGA	9E	3519
		DL	689
			715
			792
			873
			877
			1075
			1427
			1445
			1595
			1705
			1715
			1887
			1899
			2006
			2083
			2457
		EV	5031

Flights By Month



Columns: Month

Rows: CNT(nycflightinform..

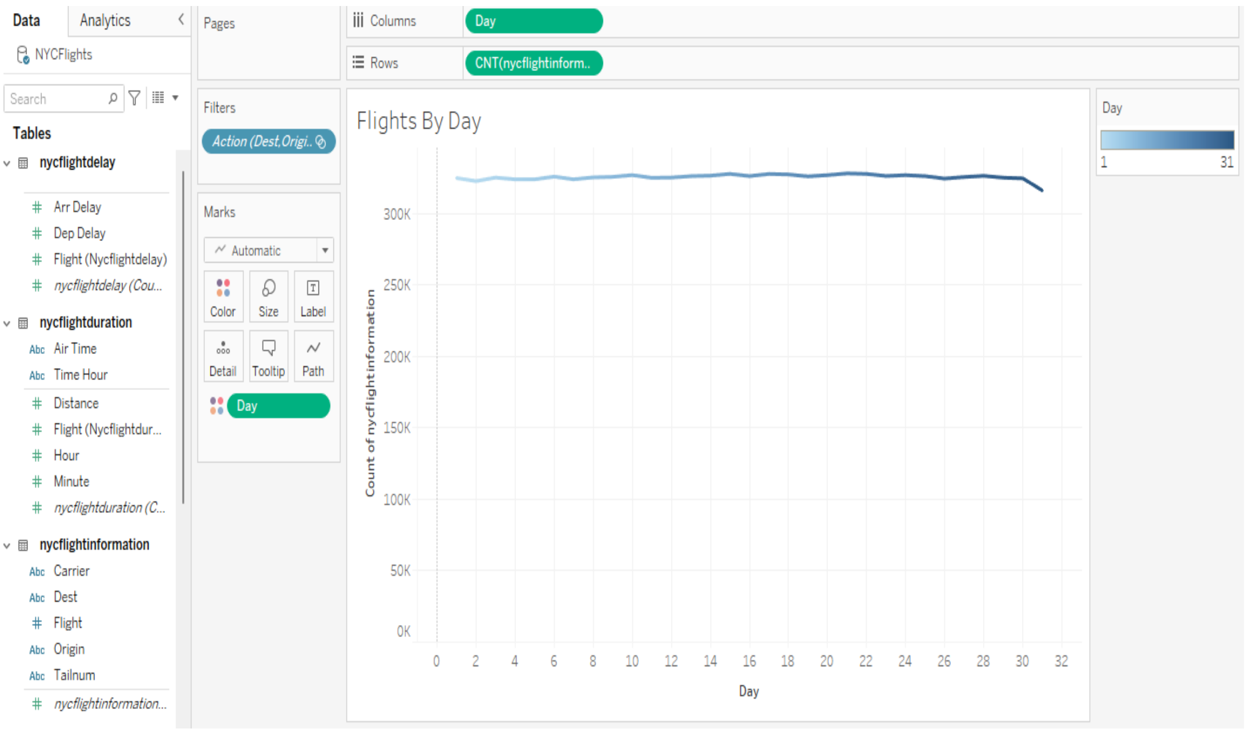
Filters: Action (Dest,Origi..

Marks: Automatic; Color: Month

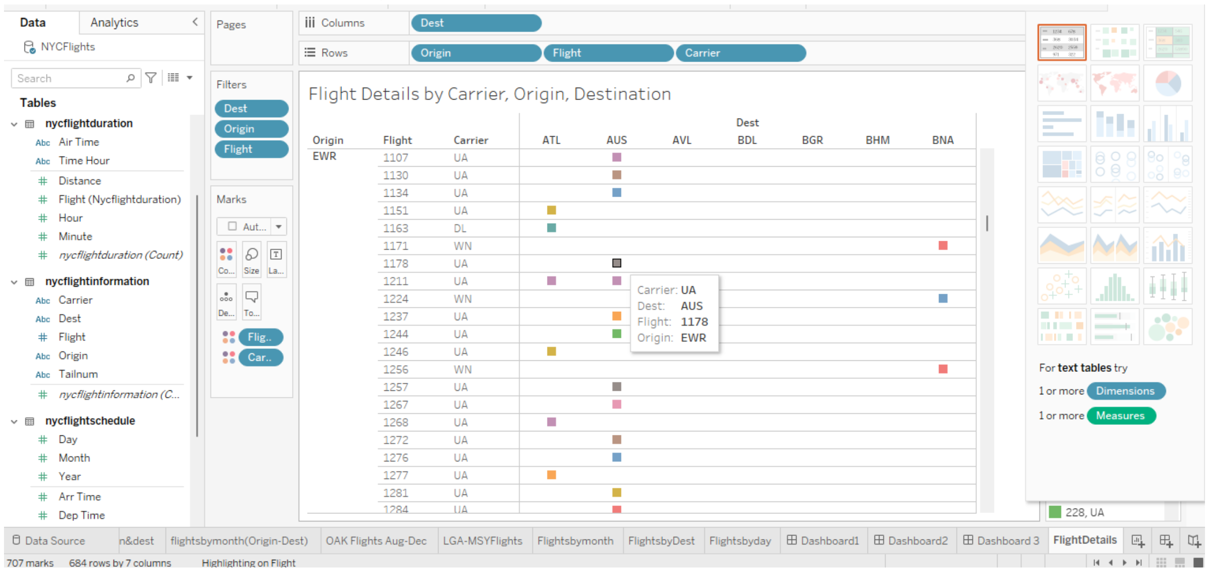
Month: 1 – 12

Flights By Month

Flights By Day



Flight Details By Carrier, Origin, Destination

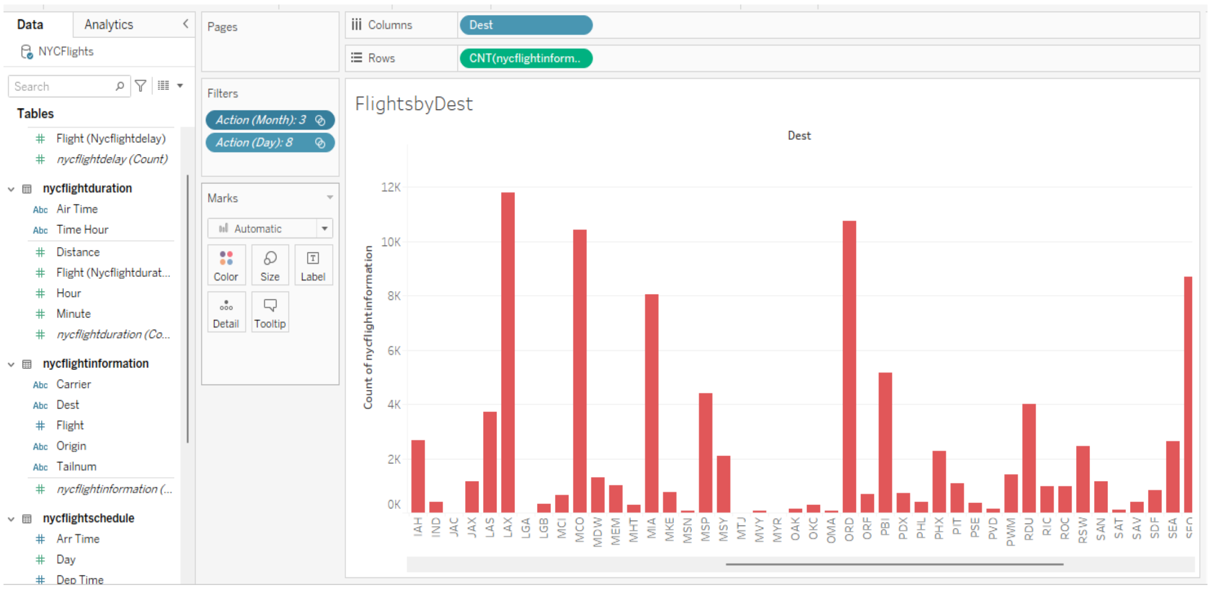


Flight Details by Carrier, Origin, Destination

			Dest						
Origin	Flight	Carrier	ATL	AUS	AVL	BDL	BGR	BHM	BNA
EWR	1107	UA		■					
	1130	UA		■					
	1134	UA		■					
	1151	UA	■						
	1163	DL	■						
	1171	WN							■
	1178	UA		■					
	1211	UA	■	■					
	1224	WN							■
	1237	UA		■					
	1244	UA		■					
	1246	UA	■						
	1256	WN							■
	1257	UA		■					
	1267	UA		■					
	1268	UA	■						
	1272	UA		■					
	1276	UA		■					
	1277	UA	■						
	1281	UA		■					
	1284	UA		■					

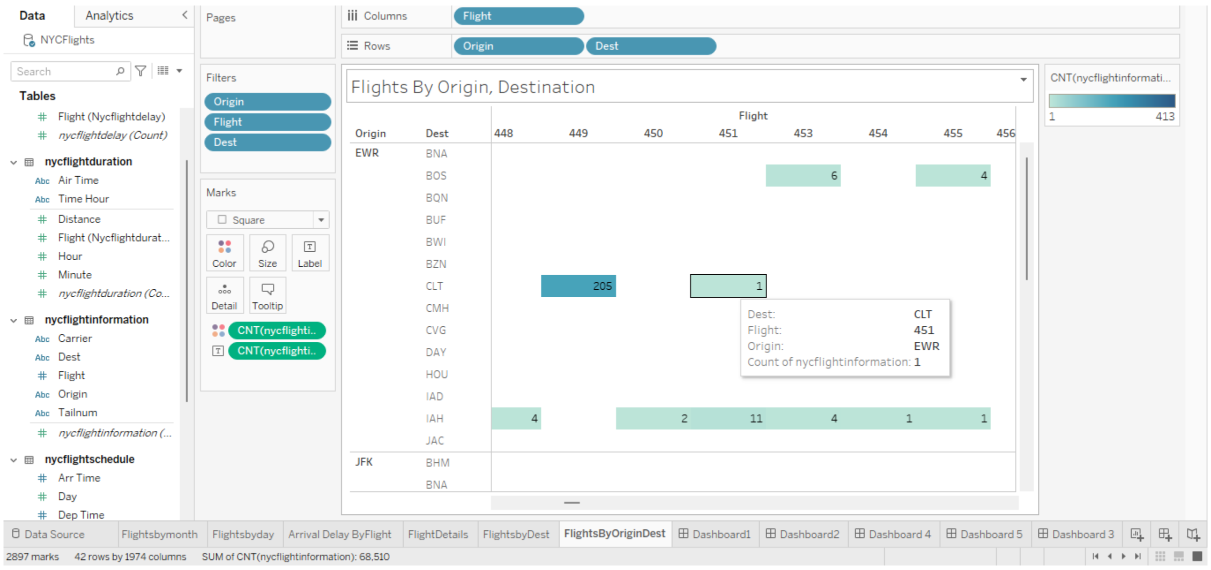
Carrier: UA
Dest: AUS
Flight: 1178
Origin: EWR

Flights By Destination



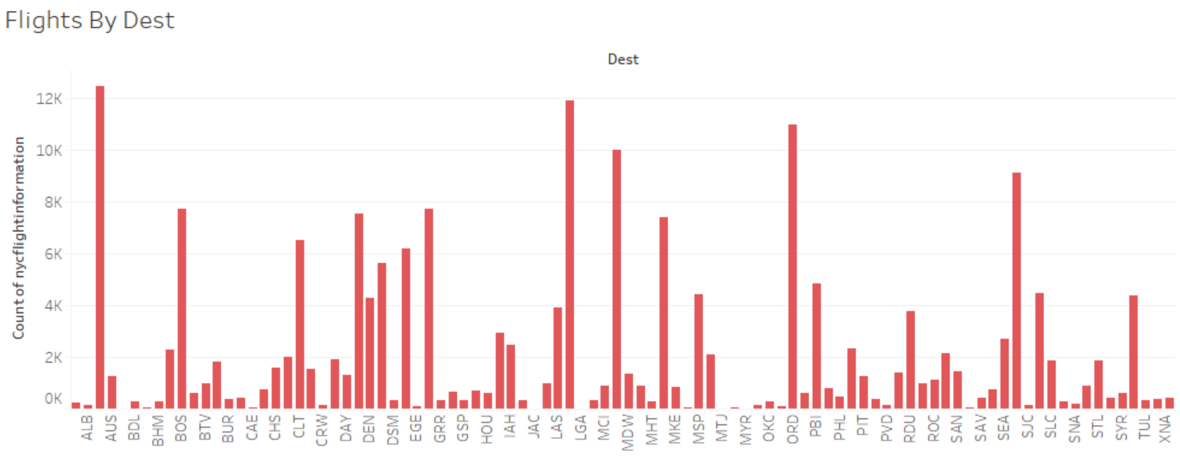
Data
Analytics
NYCFlights
Pages
Columns
Dest
Rows
CNT(nycflightinform..
Filters
Action (Month): 3
Action (Day): 8
Marks
Automatic
Color
Size
Label
Detail
Tooltip
FlightsbyDest
Dest
Count of nycflightinformation
12K
10K
8K
6K
4K
2K
0K
IAH IND JAC JAX LAS LAX LGB MCI MCO MDW MEM MHT MIA MKE MSN MSP MSY MTJ MVY MYR OAK OKC OMA ORD ORF PBI PDX PHL PHX PIT PSE PVD PWM RDU RIC ROC RSW SAN SAT SAV SDF SEA

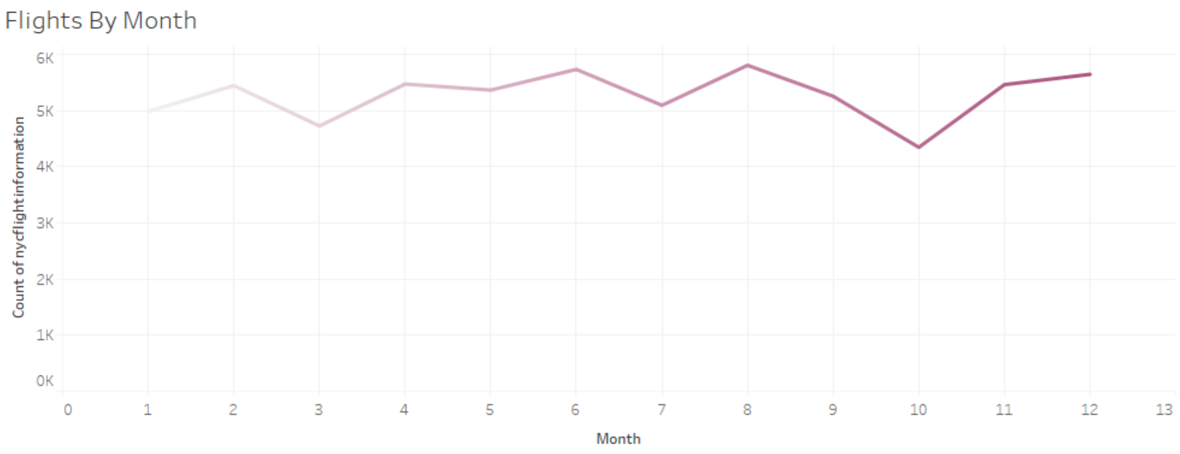
Flights by Origin, Destination



Dashboard 1:

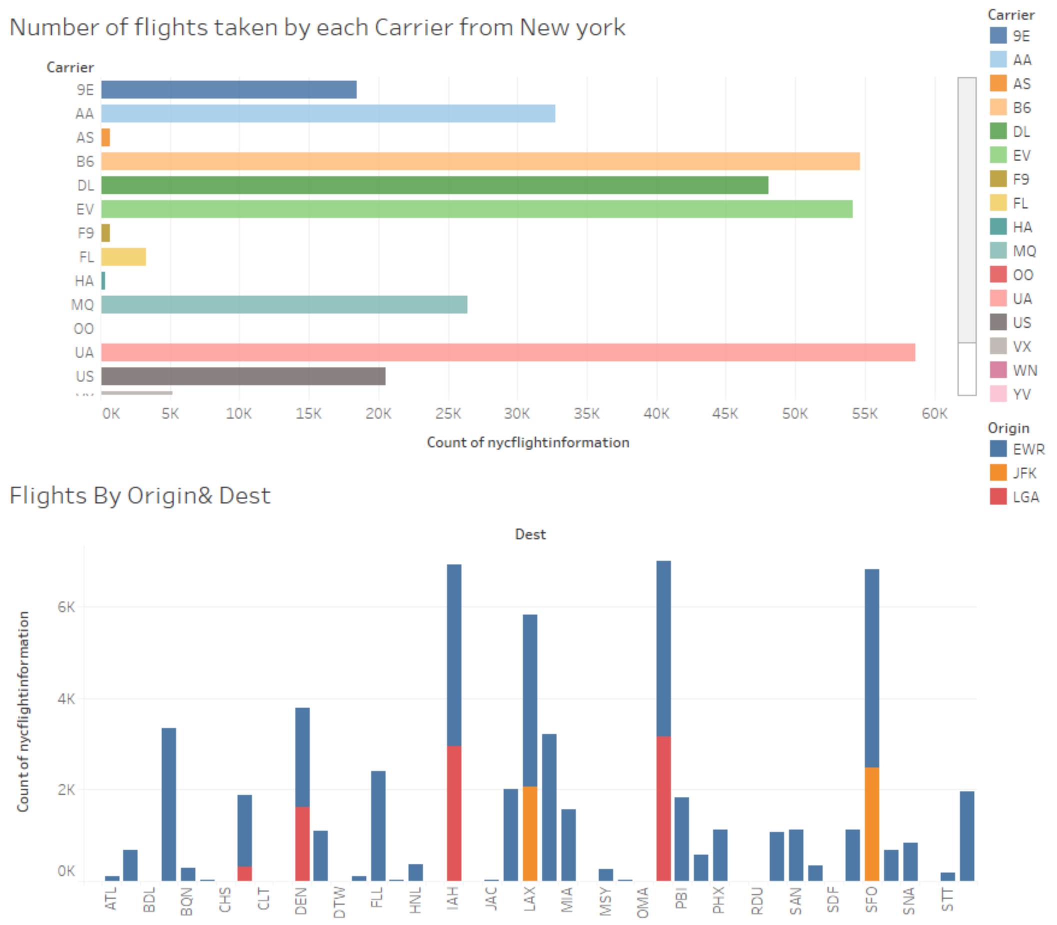
- By using this Dashboard, we can analyze the flights based on the destination and month.
- Here from the flights by destination chart, on filtering the flights by destination the number of flights by month can be updated and vice-versa.





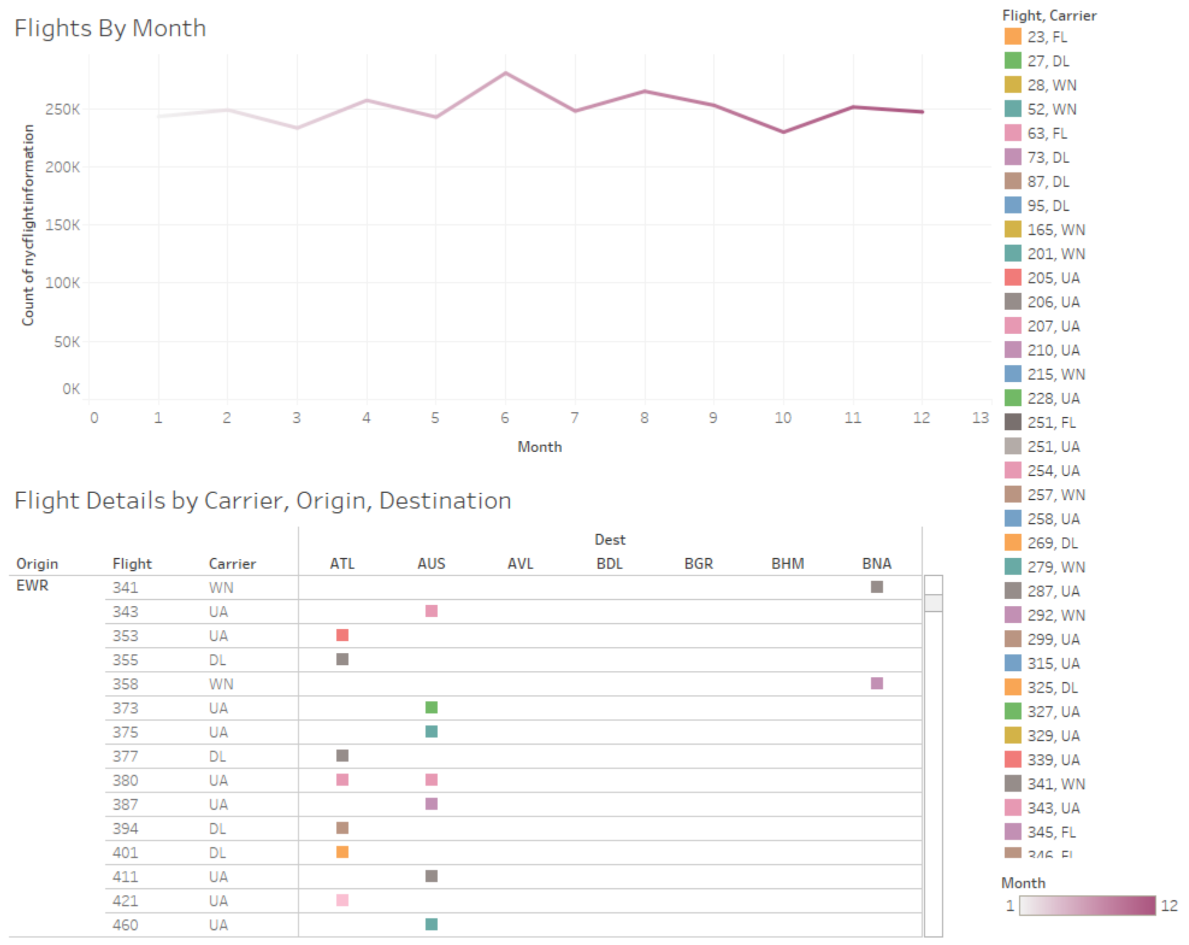
Dashboard 2:

- By using this Dashboard, we can analyze the number of flights taken by each carrier from the New York city and also flight details based on the origin and destination.
- Here from the number of flights taken by each carrier from new york chart, on filtering the flights by origin and destination the number of flights taken by each carrier can be updated and vice-versa.



Dashboard 3:

- From the below Dashboard, we can analyze the number of flights by each month and also we can get flights details by each carrier, origin and destination.
- Here from the number of flights taken by each month chart, on filtering the flights by carrier, origin and destination the number of flights taken by each month can be updated and vice-versa.

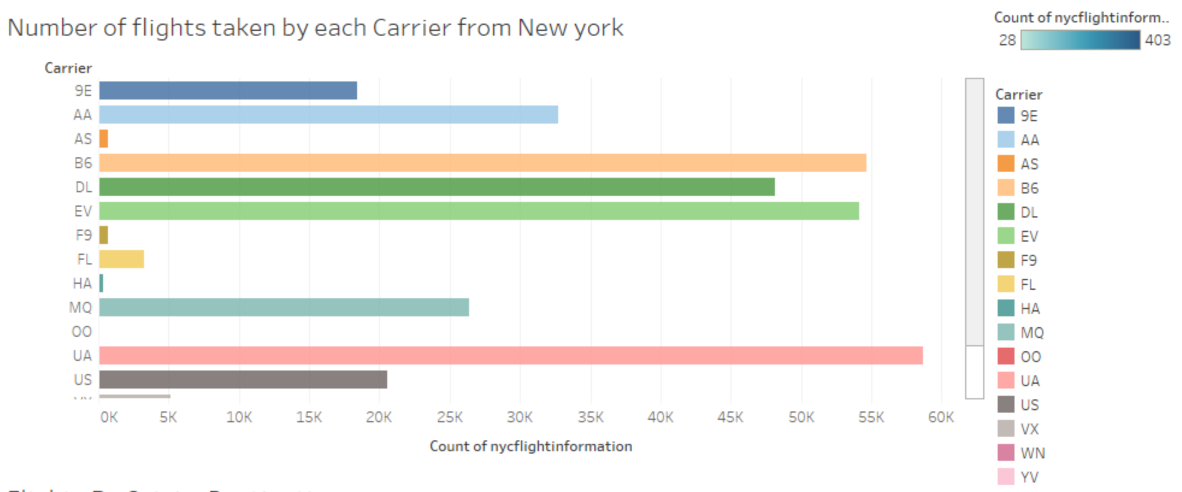

Flights By Month
Count of nycflightinformation
250K
200K
150K
100K
50K
0K
0
1
2
3
4
5
6
7
8
9
10
11
12
13
Month
Flight Details by Carrier, Origin, Destination
Month
1
12

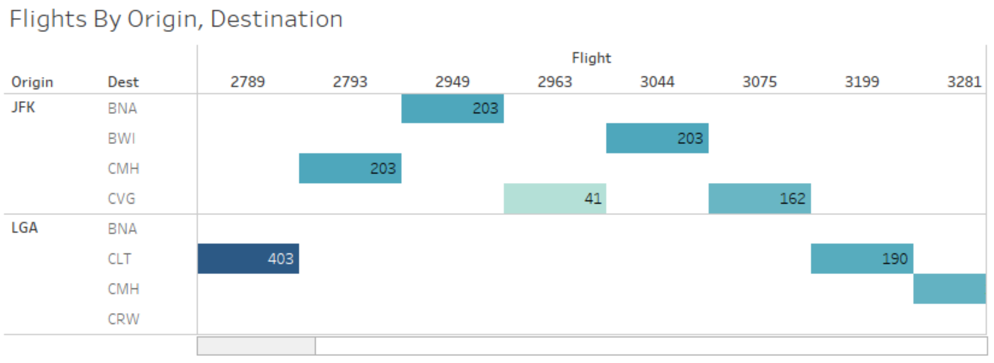
			Dest						
Origin	Flight	Carrier	ATL	AUS	AVL	BDL	BGR	BHM	BNA
EWR	341	WN							■
	343	UA		■					
	353	UA	■						
	355	DL	■						
	358	WN							■
	373	UA		■					
	375	UA		■					
	377	DL	■						
	380	UA	■	■					
	387	UA		■					
	394	DL	■						
	401	DL	■						
	411	UA		■					
	421	UA	■						
	460	UA		■					

Flight, Carrier
- ■ 23, FL
- ■ 27, DL
- ■ 28, WN
- ■ 52, WN
- ■ 63, FL
- ■ 73, DL
- ■ 87, DL
- ■ 95, DL
- ■ 165, WN
- ■ 201, WN
- ■ 205, UA
- ■ 206, UA
- ■ 207, UA
- ■ 210, UA
- ■ 215, WN
- ■ 228, UA
- ■ 251, FL
- ■ 251, UA
- ■ 254, UA
- ■ 257, WN
- ■ 258, UA
- ■ 269, DL
- ■ 279, WN
- ■ 287, UA
- ■ 292, WN
- ■ 299, UA
- ■ 315, UA
- ■ 325, DL
- ■ 327, UA
- ■ 329, UA
- ■ 339, UA
- ■ 341, WN
- ■ 343, UA
- ■ 345, FL
- ■ 346, FL

Dashboard 4:

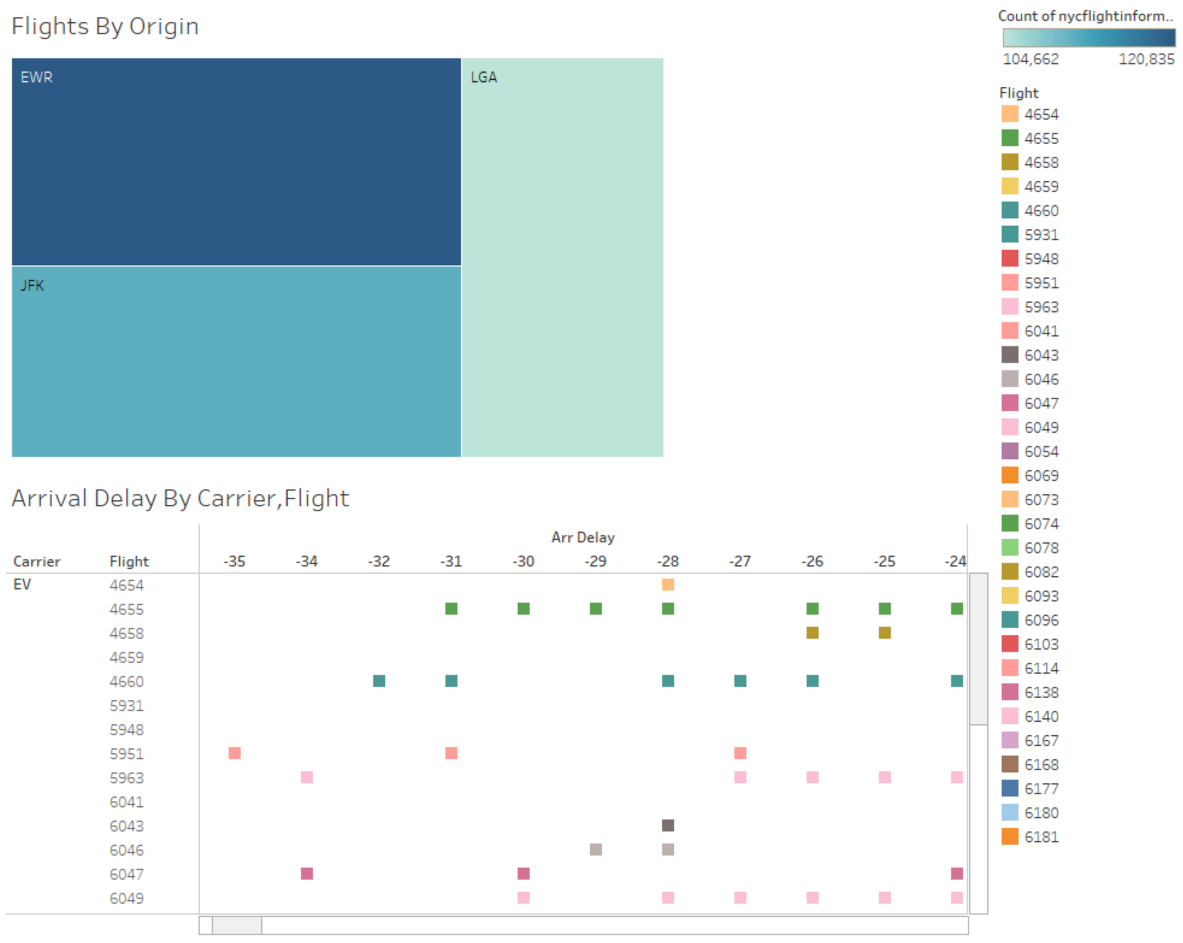
- From the below Dashboard, we can analyze the number of flights by each carrier from New York city and also we can get number of flights by origin and destination.
- Here from the number of flights taken by each carrier from new york chart, on filtering the flights by each carrier then the number of flights by origin, destination can be updated and vice-versa.





Dashboard 5:

- From the below Dashboard, we can analyze the number of flights by origin and also we can get details of arrival delay by carrier and flight.
- Here from the flights by each origin chart, on filtering the flights by each origin then the arrival delay by carrier,flight can be updated and vice-versa.



CONCLUSION:

By this Analysis, people can find the flights count and information by carrier, origin, destination by month and day. we could also analyze the arrival and departure delays so that people can avoid taking the wrong flight, avoid delays, which allows them to save time.

Airline management can improve their service by rectifying the arrival and departure delays and also by adding more details like reason for delays, ticket price. This improvement makes an advantage to the people in choosing the best flights to fly.