

CRUZALD  
ADRIANE  
VICTOR

# Brainstation : SOUTHWEST AIRLINES CO.

# KEY TERMS

- ▶ FL\_DATE – flight date
- ▶ AIRLINE – airline name
- ▶ FL\_number – flight number
- ▶ Origin\_city – city of origin for the flight
- ▶ Dest\_city – city of destination for the flight
- ▶ Dep\_delay – departure delay in minutes
- ▶ Arr\_delay – arrival delay in minutes
- ▶ Cancelled – if flight is cancelled(y/n)
- ▶ Delay\_due\_carrier – carrier delay in minutes
- ▶ Delay\_due\_weather – weather delay in minutes
- ▶ Delay\_due\_nas – national air system delay in minutes
- ▶ Delay\_due\_security – security delay in minutes
- ▶ Delay\_due\_late\_aircraft – late aircraft delay in minutes

# INTRODUCTION

- ▶ Southwest Airlines was founded 56 years ago in 1967
- ▶ Formerly known as Air Southwest
- ▶ Largest low-cost carrier or low-cost airline as of April 2020 (WIKI)
- ▶ Operates with low-cost carrier model
  - Budget, discount carrier
  - Lower fares
  - Reduced comfort

## Revenue [\[ edit \]](#)

*Forbes' world's biggest public airline companies by revenue (as of January 2023)<sup>[1]</sup>*

Rank ↕	Airline ↕	Country ↕	Revenue ↕	Profit ↕	Assets ↕	Market cap. ↕	Employees ↕
			(US\$ billions)				
1	Delta Air Lines	 United States	53.99	1.90	73.13	21.81	95,000
2	American Airlines Group	 United States	52.26	1.77	66.79	9.05	129,700
3	United Airlines Holdings	 United States	48.82	1.92	70.42	14.7	92,800
4	Lufthansa Group	 Germany	35.8	0.94	48.79	12.10	109,509
5	Air France–KLM	 France/  Netherlands	29.41	0.97	34.81	4.12	78,950
6	Southwest Airlines	 United States	24.83	0.65	35.55	17.65	66,000
7	International Airlines Group	 UK/  Spain	24.28	1.15	41.95	9.41	66,044
8	Turkish Airlines	 Turkey	19.68	2.79	32.00	8.88	40,264
9	China Southern Airlines	 China	12.92	−4.81	44.92	12.54	100,431
10	Air Canada	Canada	12.72	−1.31	21.78	5.33	42,916

# HYPOTHESIS

- ▶ Number of Cancellations are directly proportional to number of flights
- ▶ Did Covid affect Southwest Airlines' operations?
  - COVID WAS DETECTED LATE 2019, ([webmd.com](#))
  - TRAVEL BAN WAS IMPLEMENTED MARCH 2020 BY TRUMP ADMINISTRATION, ([webmd.com](#))
  - BAN LIFTED JUNE 2022 ([theamericandream](#))
- ▶ What preventative measures or policies do other companies for saving money have that Southwest airlines don't?

# DATA PROCESSING

DOWNLOAD .CSV DATASET FROM KAGGLE:

- [Flight Delay and Cancellation Dataset \(2019-2023\) \(kaggle.com\)](https://www.kaggle.com/datasets/bureauoftransportationstatistics/flight-delay-and-cancellation-dataset-2019-2023)
- ORIGINALLY FROM BUREAU OF TRANSPORTATION STATISTICS
- PRE-MERGED IN KAGGLE FROM 2019 JANUARY TO 2023 AUGUST
- VARIABLES INCLUDE ORIGIN, DESTINATION, CANCELLATION REASONS, AND DELAYS IN MINUTES(VARYING).

# DATA PROCESSING

CREATE TABLE IN MYSQL WORKBENCH NAMED 'flightdata'

Column	Type	Default Value	Nullable	Character Set	Collation	Privileges	Extra
✧ fl_date	varchar(45)		YES	utf8mb4	utf8mb4_0900_...	select,insert,update,references	
✧ airline	varchar(45)		YES	utf8mb4	utf8mb4_0900_...	select,insert,update,references	
✧ airline_dot	varchar(45)		YES	utf8mb4	utf8mb4_0900_...	select,insert,update,references	
✧ airlinecode	varchar(45)		YES	utf8mb4	utf8mb4_0900_...	select,insert,update,references	
✧ dot_code	varchar(45)		YES	utf8mb4	utf8mb4_0900_...	select,insert,update,references	
✧ fl_number	varchar(45)		YES	utf8mb4	utf8mb4_0900_...	select,insert,update,references	
✧ origin	varchar(45)		YES	utf8mb4	utf8mb4_0900_...	select,insert,update,references	
✧ origin_city	varchar(45)		YES	utf8mb4	utf8mb4_0900_...	select,insert,update,references	
✧ dest	varchar(45)		YES	utf8mb4	utf8mb4_0900_...	select,insert,update,references	
✧ dest_city	varchar(45)		YES	utf8mb4	utf8mb4_0900_...	select,insert,update,references	
✧ crs_dep_time	varchar(45)		YES	utf8mb4	utf8mb4_0900_...	select,insert,update,references	
✧ dep_time	varchar(45)		YES	utf8mb4	utf8mb4_0900_...	select,insert,update,references	
✧ dep_delay	varchar(45)		YES	utf8mb4	utf8mb4_0900_...	select,insert,update,references	
✧ taxi_out	varchar(45)		YES	utf8mb4	utf8mb4_0900_...	select,insert,update,references	
✧ wheels_off	varchar(45)		YES	utf8mb4	utf8mb4_0900_...	select,insert,update,references	
✧ wheels_on	varchar(45)		YES	utf8mb4	utf8mb4_0900_...	select,insert,update,references	
✧ taxi_in	varchar(45)		YES	utf8mb4	utf8mb4_0900_...	select,insert,update,references	
✧ crs_arr_time	varchar(45)		YES	utf8mb4	utf8mb4_0900_...	select,insert,update,references	
✧ arr_time	varchar(45)		YES	utf8mb4	utf8mb4_0900_...	select,insert,update,references	
✧ arr_delay	varchar(45)		YES	utf8mb4	utf8mb4_0900_...	select,insert,update,references	
✧ cancelled	varchar(45)		YES	utf8mb4	utf8mb4_0900_...	select,insert,update,references	
✧ cancelled_code	varchar(45)		YES	utf8mb4	utf8mb4_0900_...	select,insert,update,references	
✧ diverted	varchar(45)		YES	utf8mb4	utf8mb4_0900_...	select,insert,update,references	
✧ crs_elapsed_time	varchar(45)		YES	utf8mb4	utf8mb4_0900_...	select,insert,update,references	
✧ elapsed_time	varchar(45)		YES	utf8mb4	utf8mb4_0900_...	select,insert,update,references	
✧ air_time	varchar(45)		YES	utf8mb4	utf8mb4_0900_...	select,insert,update,references	
✧ distance	varchar(45)		YES	utf8mb4	utf8mb4_0900_...	select,insert,update,references	
✧ delay_due_carrier	varchar(45)		YES	utf8mb4	utf8mb4_0900_...	select,insert,update,references	
✧ delay_due_weather	varchar(45)		YES	utf8mb4	utf8mb4_0900_...	select,insert,update,references	
✧ delay_due_nas	varchar(45)		YES	utf8mb4	utf8mb4_0900_...	select,insert,update,references	
✧ delay_due_security	varchar(45)		YES	utf8mb4	utf8mb4_0900_...	select,insert,update,references	
✧ delay_due_late_aircraft	varchar(45)		YES	utf8mb4	utf8mb4_0900_...	select,insert,update,references	

# DATA PROCESSING

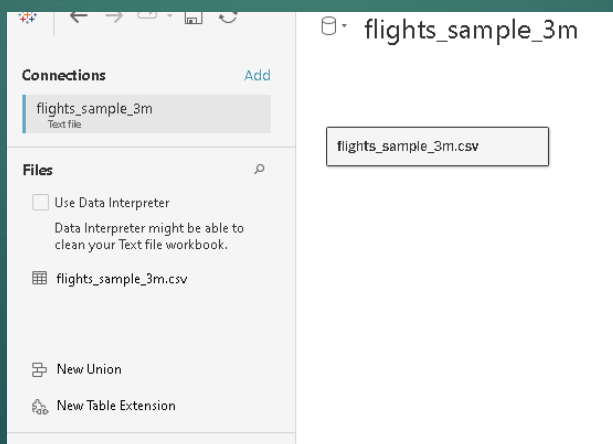
## ► LOAD .CSV TO MYSQL USING COMMAND:

```
load data infile 'C:/Users/cruza/OneDrive/Desktop/BrainStation Files/  
Brainstation FINAL PROJECT/Flight Cancellation Dataset/flights_sample_3m.csv'  
ignore into table flightdata  
FIELDS TERMINATED BY ','  
ENCLOSED BY '"' LINES TERMINATED BY '\n' IGNORE 1 rows;
```

## ► RESULT:

✓	10	15:55:47	SHOW VARIABLES LIKE "secure_file_priv"	1 row(s) returned
✗	11	15:56:07	load data infile 'C:/Users/cruza/OneDrive/...	Error Code: 29. File 'C:/ProgramData/MySQL/MySQL Server 8.0/Data/UserscruzaOneDriveDesk
✓	12	15:57:08	load data infile 'C:/Users/cruza/OneDrive/...	3000000 row(s) affected Records: 3000000 Deleted: 0 Skipped: 0 Warnings: 0

## ► LOAD .CSV TO TABLAEU



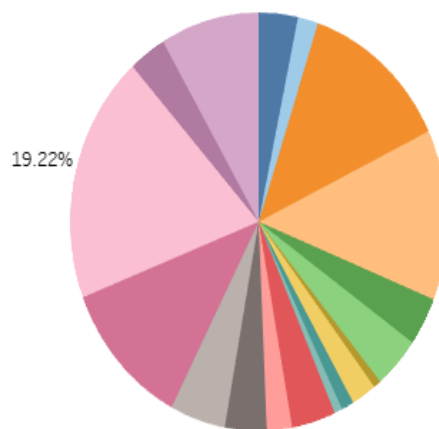
# DETERMINING OUTLIERS OF DELAYS AND CANCELLATIONS

```
#show airline flights between 2019 and 2023#
```

```
• Select airline, count(airline) AS flightCount from flightdata group by airline order by flightCount desc;
```

	airline	flightCount
►	Southwest Airlines Co.	576470
	Delta Air Lines Inc.	395239
	American Airlines Inc.	383106
	SkyWest Airlines Inc.	343737
	United Air Lines Inc.	254504
	Republic Airline	143107
	Envoy Air	121256
	JetBlue Airways	112844
	Endeavor Air Inc.	112463
	PSA Airlines Inc.	107050
	Alaska Airlines Inc.	100467
	Spirit Air Lines	95711
	Mesa Airlines Inc.	65012
	Frontier Airlines Inc.	64466
	Allegiant Air	52738
	Hawaiian Airlines Inc.	32114

AIRLINE PIE



## Data Summary

This viz

- Has 18 data points

## Detected Outliers

Tableau detected these marks as unusual compared to other marks in the selected viz.

**High** Southwest Airlines Co.  
Count of flights\_sample\_3m.csv: 576,470

**High** Delta Air Lines Inc.  
Count of flights\_sample\_3m.csv: 395,239



# TALLY OF DELAYS/CANCELLATIONS

```
select airline,
count(
case
#when cancelled = 1 THEN 1
when dep_delay > 0 THEN 1
when arr_delay > 0 THEN 1
when DELAY_DUE_CARRIER > 0 THEN 1
WHEN DELAY_DUE_WEATHER > 0 THEN 1
WHEN DELAY_DUE_NAS > 0 THEN 1
WHEN DELAY_DUE_SECURITY > 0 THEN 1
WHEN DELAY_DUE_LATE_AIRCRAFT > 0 THEN 1
else 0
end ) as Problematic_Flight_Count
from flightdata
where airline = "Southwest Airlines Co.";
```

Result Grid									
Filter Rows:									
Export: Wrap Cell Content:									
airline	DEP_DELAY	ARR_DELAY	DELAY_DUE_CARRIER	DELAY_DUE_WEATHER	DELAY_DUE_NAS	delay_due_security	delay_due_late_aircraft	CANCELLED	
Southwest Airlines Co.	360663	274799	96998	3401	56458	693	98540	26419	

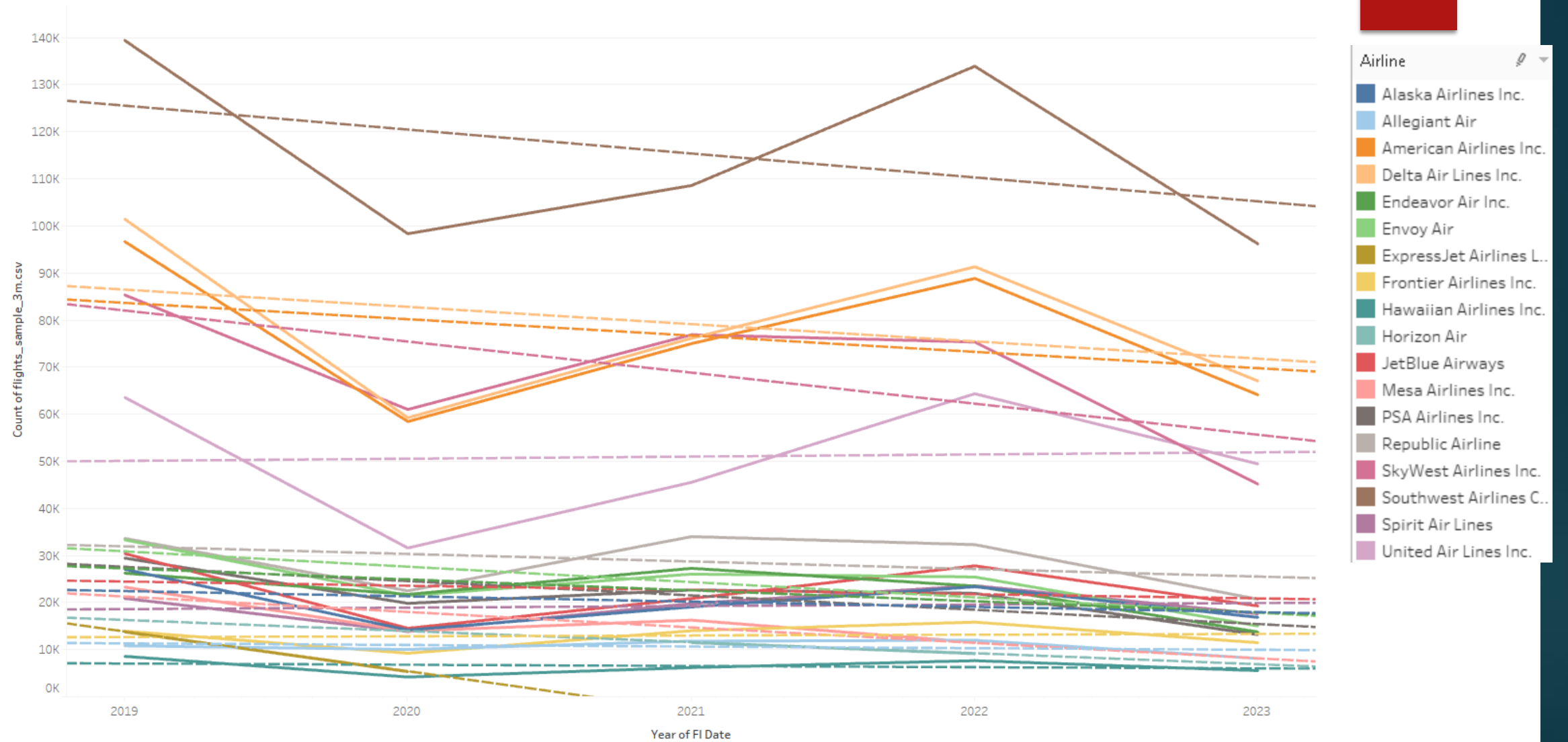
# FLIGHT DELAYS/CANCELLATIONS PER YEAR

```
#Show date counts per year
```

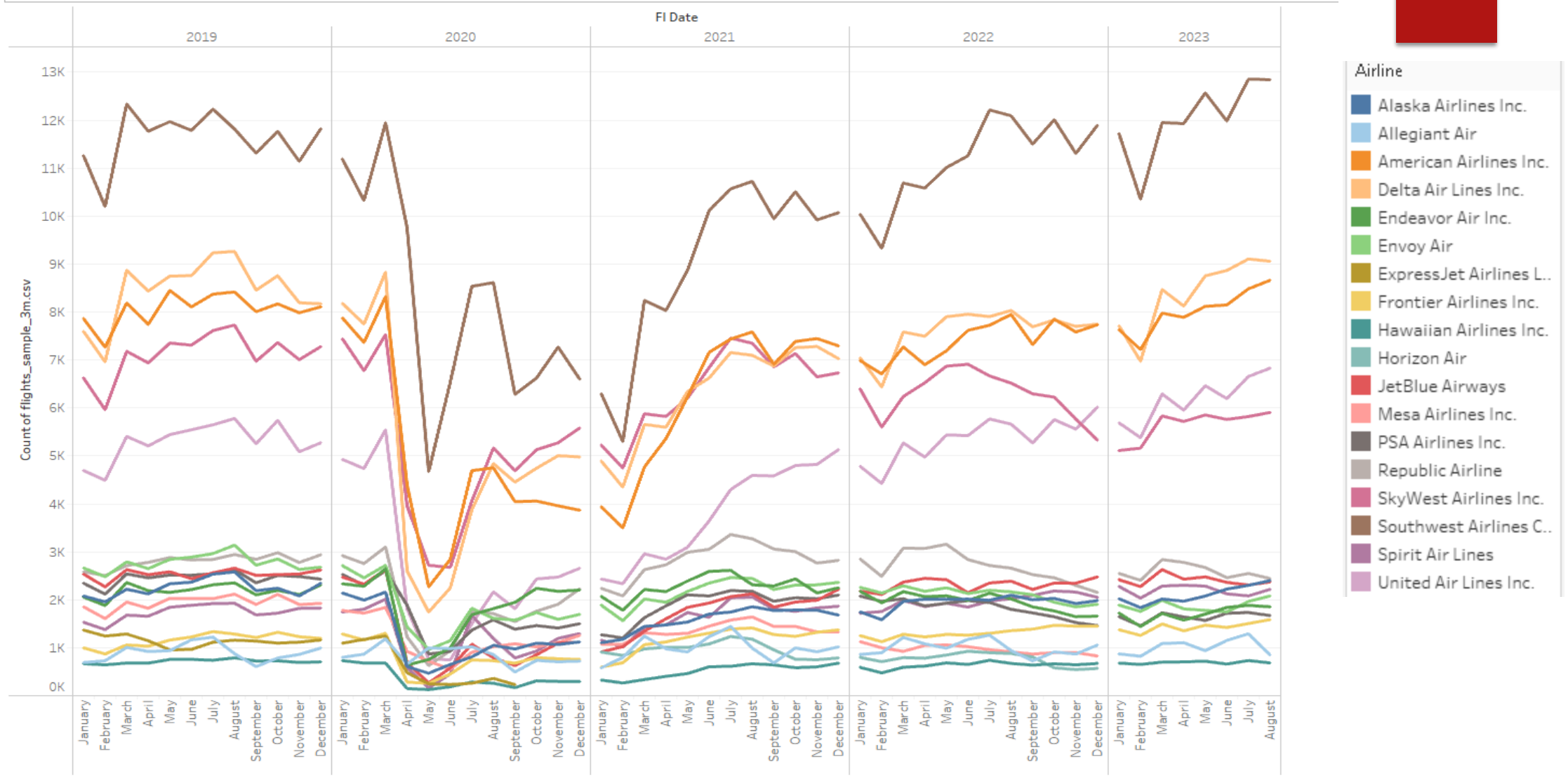
```
select substring(fl_date,1,4) as flYear, count(fl_date) as Cnt from flightdata  
where airline = "Southwest Airlines Co." group by substring(fl_date,1,4);
```

	flYear	Cnt
►	2019	139406
	2021	108595
	2020	98346
	2022	133923
	2023	96200

## TRENDS



## 2019 to 2023 Cancellation by Month



# CITIES WITH THE MOST FLIGHT DELAYS/CANCELLATIONS

## ► ORIGIN CITY

```
#show denver and chicago as percentage of flights both origin and destination
• select origin_city, (count(origin_city)/(select count(origin_city) from flightdata
where airline = "Southwest Airlines Co.))*100 AS originCityPerc from flightdata where airline = "Southwest Airlines Co."
group by origin_city order by originCityPerc desc LIMIT 5;
```

	origin_city	originCityCount
►	Denver, CO	36420
	Chicago, IL	35260
	Las Vegas, NV	31305
	Baltimore, MD	29368
	Dallas, TX	29257

## ► DESTINATION CITY

```
#show denver and chicago as percentage of flights destination
• select dest_city, (count(dest_city)/(select count(dest_city) from flightdata
where airline = "Southwest Airlines Co.))*100 AS destCityPerc from flightdata where airline = "Southwest Airlines Co."
group by dest_city order by destCityPerc desc LIMIT 5;
```

	dest_city	destCityCount
►	Denver, CO	36446
	Chicago, IL	34962
	Las Vegas, NV	31641
	Baltimore, MD	29632
	Dallas, TX	29142

# DERIVED INSIGHTS

NUMBER OF CANCELLATIONS AND DELAYS WERE AFFECTED BY THE START OF TRAVEL BAN IN 2020

THERE IS A NEGATIVE TREND OF CANCELLATIONS AND FLIGHT DELAYS FROM 2019 TO 2023

DENVER AND CHICAGO ARE THE CITIES MOST TRAVELLED TO AND FROM THAT HAVE DELAYS/CANCELLATIONS

ACCORDING TO THE CIRIUM, ON-TIME PERFORMANCE REVIEW 2022 REPORT, SOUTH-WEST SCORED AN OTP OF 74.06%, MAKING IT ONE OF THE TOP 5 MAINLINE AIRLINES IN THE N.A. REGION

# CONCLUSION

COVID AFFECTED THE NUMBER OF FLIGHTS, BUT IMPROVED THE NUMBER OF CANCELLATIONS



SOUTHWEST AIRLINES IS LOSING REVENUE (6TH IN REVENUE AS OF JANUARY 2023) ACCORDING TO WIKI



STILL TOP 5 N.A. AIRLINE ACCORDING TO 2022 REPORT



SOUTHWEST AIRLINES HAS THE MOST NUMBER OF FLIGHTS FLOWN COMPARED TO OTHER AIRLINE COMPANIES

# RECOMMENDATIONS

IMPROVE COMPANY  
POLICY BY CHARGING  
PENALTIES ON  
CUSTOMERS

IMPROVE AIRLINE  
OPERATIONS AND  
MAINTENANCE  
POLICIES

BETTER SCHEDULING  
PRACTICES

ADVANCED  
PLANNING

COLLABORATE/SHARE  
DATA WITH OTHER  
COMPANIES TO  
IMPROVE ON-TIME  
PERFORMANCE(OTP)



# REFERENCES:

1. U.S. Department of Transportation - Air Travel Consumer Report: <https://www.transportation.gov/airconsumer/air-travel-consumer-reports>
2. FlightStats - Global Flight Tracker and Travel Planning: <https://www.flightstats.com/>
3. Flight Delays – What can be improved?: [\(25\) Flight Delays: What can be improved? | LinkedIn](#)
4. [U.S. Passenger Carrier Delay Costs | Airlines For America](#)
5. [Coronavirus History: How Did Coronavirus Start? \(webmd.com\)](#)
6. [End of the US Travel Ban \(the-american-dream.com\)](#)
7. [Largest airlines in the world - Wikipedia](#)