U.S Airline Time Study, On-Time Statistics and Delay Causes

The complete tableau workbook has been published in this link
https://public.tableau.com/profile/libin.guo#!/vizhome/U_SFightsOnTimePerformance/U_SFlightStatiticOntimeperformanceandDelayCause?publish=yes

The initial project without change from feedback is published in this link
https://public.tableau.com/profile/libin.guo#!/vizhome/InitialU_SFightsProject/U_SFlightStatiticOntimeperformanceandDelayCause?publish=yes

- ## Summary

The aim of this project is to analyze U.S Flight Geographical distribution,investigate the performance of flights over time and dive deeper into Delay Causes and each airline performance using Tableau. There are two datasets used in this analysis- Flight statistic from 2003 to 2017 (US Government website) and airport dataset. There are 10 charts, 2 dashboard and 1 story created in this project to show the overall trend of U.S Operation Flight Number for airport nationwide on the map, the flight on time rate,delay rate and cancelled rate by year and by carrier, along with the delay cause analysis -non weather cause delay vs weather cause delay.

- ## Design

Part 1- Introduction of Dataset

The U.S. Department of Transportation's (DOT) Bureau of Transportation Statistics (BTS) tracks the on-time performance of domestic flights operated by large air carriers.This flight On-Time Performance was downloaded from The U.S. Department of Transportation's (DOT) Bureau website http://stat-computing.org/dataexpo/2009/the-data.html. The table dictionary can be found from this link https://www.transtats.bts.gov/TableInfo.asp. This table contains on-time arrival data for non-stop flights by major air carriers, and provides such additional items as departure and arrival delays, origin and destination airports, flight numbers, scheduled and actual departure and arrival times, cancelled or diverted flights, taxi-out and taxi-in times, air time, and non-stop distance.

There is 232,633 records and 21 columns in the dataset. The time frame for this dataset is from 2003 to 2017.
I also downloaded US airport data from http://ourairports.com/data/ to join with flight data so I can create map visualization.

Part 2- Data Wrangling

At first, the type of some variable should be changed.

- Month and Year variable should be combined to create a date calculated field
- Fight Number should string instead of numeric
- Airport variable should be assigned geographical role as airport

Calculated fields:

Flight Operation: [Arr Cancelled]+[Arr Diverted]+[Arr Del15]+[Arr Flights]

(Note: Arr_del15 is the sum of carrier_ct, weather_ct, nas_ct, security_ct and late aircraft)

On Time Percentage: sum([Arr Flights])/sum(([Arr Flights]+[Arr Cancelled] + [Arr Del15]+ [Arr Diverted]))

Delayed(%): sum([Arr Del15])/sum([Flight Operation])

Cancelled & Diverted(%): sum(([Arr Diverted]+[Arr Cancelled]))/sum([Flight Operation])

On Time performance Group: IIF([On Time Percentage] >0.88,'Excellent',IIF([On Time Percentage]>0.82, 'Good','Not Good'))

Non-weather delay: [Security Delay]+[Carrier Delay]+[Nas Delay]+[Late Aircraft Delay]

Non weather Delay/weather Delay Ratio: sum([Non-weather Delay])/sum([Weather Delay])

Part 3- Visualization

I created 10 charts, 2 dashboard and 1 story in this project.

1. **U.S Flight trend from 2003 and 2017**

The first chart "Number of Fights from 2003 -2017" is line chart to show the trend of Flight number, from 2003 to 2017, the total number of operation flight increase from 2003 and stopped at 2007, which is the peak in past 14 years, then kept decreasing from 2007.If we looked more carefully about cycle of trend in the chart, we can see that there are a lot peaks of higher number of flights occurred in summer time(June, July and August).

I am also interested to see which city/state have highest number of flights and busiest airport overall, hence, the second chart is horizontal bar chart --"Municipality Fight Distribution".From this chart, we can see Chicago's total number of flight is the Top 1 in U.S.A,then the Atlanta, Dallas-Fort Worth, Houston and new York.

2. **Fights On time performance**

On time performance is the focus of this project, "On Time Percentage Over Time" and "Delay, Cancelled & Diverted Flights Rate" all use line chart to examine the trend of on time performance over past 14 years in the same dashboard. We can also see that the average on time flight percentage is between 80% to 90% over time. 2012 Q4 has the highest flight on time rate. On the other hand, 2008 and 2004 is the two peak of Delayed Rate, overall, both delayed rate and cancelled & diverted rate are decreasing .

The next chart is “OnTimePercnt by Carrier Line Chart”, which can allow us to see which carrier /airline have highest on time rate and being consistent over the years. The result shows that the Hawaiian Airlines and Alaska Airlines have best on time flight percentage which is much higher than other airline and being stable over the years,. Also, I did not see big airline company stand out in this chart.
Also, i am curious to see which airport has excellent, good and not good on time percentage, therefore, i created “U.S Large Airport On Time Performance Map” to filter large airport and group airport based on the performance as following and and assign them different color displayed in legend.
On time percentage >=88% : Excellent
88%> On time percentage>82%: Good
82% > On time percentage : Not Good

Furthermore, since I live in Dallas, TX and traveled by flight regularly, I want to see the carrier’s performance on DFW area. The chart shows Atlantic Southeast Airline has the worst delay rate, which is almost 50%. American Airlines and Delta actually has best flight performance in DFW area.

3. Flight Delay Cause

Last part of analysis focus is Delay Cause, I created “Delay Cause by Year in Total Delay Minutes Bar Chart” to show the what’s reason caused the delay of flight and its trend over times. The charts shows that Late Aircraft Delay, carrier Delay, NAS Delay ,Weather Delay and security delay are 5 main cause of flight delay. Among them, Late AirCraft Delay is the number 1 reason for flight delay . Over times, all delay causes’ number of flights is decreasing.

A dashboard with “Non weather Delay vs Weather Delay Ratio Trend from 2003 to 2017 Bar Chart” and “Carrier Non Weather Delay and Weather Delay Ratio Bar Chart” is created to showcase the Non weather Delay vs Weather Delay Flight trends over years and to check if weather caused delay reduced. On on time performance dashboard, we see the delay rate is decreasing over time.
Based on calculated field - Non weather and weather delay ratio, the ratio is actually slightly increasing over the years. 2011 and 2014 are two peaks of the trend. Aloha, ATA and AirTrans have the highest Non Weather vs weather delay ratio. For Aloha airline, it does not have any flights after 2008 Q1, so the statistic for Aloha is not useful. However, we can conclude over years, weather caused delay is reduced.

- **Feedback**

After I created the 1st chart- “Number of Operation Fights from 2003 -2017”, the feedback has be provided by my friends.

1. The first chart could be shortened by excluding the partial data from 2003. Starting with Q3 would give a more accurate year to year comparison and increased detail between quarters.

Action: The original chart is a little misleading- number in 2003 has a huge jump to 2014. I changed the date from quarter year to month year since 2003 data is not complete and only start from June 2003.Also, y-axis is narrowed to remove the space on the chart.

2. The text boxes are too small both in comparison to the charts and individually they require scrolling to read the full narrative.

Action: I adjusted the height of boxes and make them more use-friendly.

3. OnTimePercentage by Time (chart 3) needs a clearer title.

Action: The time has been changed to “Flight On time Rate By Year”

4. Chart 4 is too busy to be of use.

Action: I add filter on carrier and only leave 4 carrier for comparison on the charts in order to make it cleaner.

5. Chart 5 (the map) is very good.
6. Delay Cause by Year narrative needs explanation of what a “NAS Delay” is.

Action : Add whole name for NAS(National Aviation System) in Chart.

7. Non Weather Delay vs Weather Delay Ratio by carrier chart could be created using actual weather delay number, sorted by total number of delays or flights to show values for the major carriers, giving more applicable information to the average flier.

Action: The chart is recreated as Non Weather Delay vs Weather Delay By carrier Bar Chart and sorted descending number of non weather delay instead of using ratio.

- **Resources**

https://www.transtats.bts.gov/OT_Delay/OT_DelayCause1.asp