Udacity Tableau Project

Questions stated:

1. Which airlines have most delays/cancellations?
   https://public.tableau.com/profile/hanna.kondrashova#!/vizhome/UdacityTest_3/Commentswhichairlineshavemostdelayscancellations
   Design: Two figures are present which answer two parts of this question: 'Avg. delays per airline' and 'Cancelled flights per airline'. The first has a discrete (categorical) variable, airline name along the x axis, and (continuous) average total delay per airline in minutes. The bar diagram is selected due to the fact that one of the variables is categorical. To better understand the nature of total value, each bar is split into sectors colored with seven different colors ('Delay type per airline' legend). The second figure shows the bubble chart showing the percentage of cancellations per each airline, so that they better (than, for example, pie diagram) visualize the percentage with the help of the size of the bubble. The 'bubbles' are colored with respect to the airline name for better visual support.
   Summary: Total average delay is the largest for Frontier Airlines and Spirit Air Lines. The departure and average terms prevail in the total average delay. Cancellation percentage is near for most of the airlines (the largest is 5.41%, the next one is 2.80%).

2. Why the flights are delayed/cancelled?
   https://public.tableau.com/profile/hanna.kondrashova#!/vizhome/UdacityTest_3/Commentswhytheflightsaredelayedcancelled

   Design: Two figures are present: 'Avg. total delay per month' and 'Cancellation reasons'. The first one has two y-axes for average total delay in minutes, and average total delay per cause in minutes. This was motivated by the goal to show both total and the terms of it, but total turned out to be larger than any of its components, so it's more readable at a separate axis. The cancellation reason is the x-axis of the second plot, and as far as it is categorical variable, this motivates to use a bar chart. 'Security' reason has so small percentage that the labels above each bar was used to show the value. Also you can see how the cancellation is influenced by the month of the year applying the filter.

   Summary: The average total delay peaks during the holiday period in June and December. It is summed mostly from departure and arrival delay. The third most important component is late aircraft delay which also varies during the year and seems to be connected to the first two. Air system delay, which varies less during the year, has about twice less influence. For cancellations, weather has the most influence, though for April, June and October airline and weather cause very near amount of cancellations.

3. At which airports you have to wait longer for departure?
   https://public.tableau.com/profile/hanna.kondrashova#!/vizhome/UdacityTest_3/Commentatwhichairportsyouhavetowaitlongerfordeparture
   Design: Two figures are available. First shows average departure delay by total departure delay in minutes, each point on the figure has a label with the name of the airport. Both variables are real numbers (total departure delay is integer). The idea was to see how the average departure delay deviates for different airports and to see if it is connected to the total departure delay for all flights. Another figure shows the map of the US with a filter for the state and a slider for

average departure delay to show at which airports the customers have to wait longer (in average), and also to be able to focus on big delays. The airports are highlighted with different colors and depicted with circles of different size proportional to the delay.

Summary: Most of airports tend to have departure delay time of about 30 minutes. The sum of the departure delays per airport doesn't seem to be much related to the average delay, which mostly affects the customers. Airports with small total departure delay (like for Valdez Airport) can still have exceptional average delay (95.43 minutes in this case).

Preprocessing:

1. For delays the negative values were discarded with IIF function,
2. A, B, C, D for cancellation reason was converted into actual reason with IF … THEN block,
3. Inner join on airline code (and 'airline' field, respectively) for flights.csv and airlines.cvs was made to acquire human-readable representation of the airlines,
4. The same was done for origin airport for correspondent fields.

Resources:

1. Tableau Tutorial https://www.tableau.com/learn/tutorials