Summarizing Data Using Tidyverse: Alaska Airlines Delays in Departure and Arrival

AUTHOR
Rhys, Shields, Brendon, Parth

PUBLISHED
September 25, 2024

Key Findings and Summary Statistics

Alaska Airlines is a mid-size airline in the commercial aviation industry, that focuses on short international travel and regionalized domestic travel within the U.S. We found that Alaska Airlines averages about a 16.7-minute delay on departure and arrivals vs 28.46-minute delay on departure and arrivals for the commercial aviation industry. 10,208 flights were delayed in June, 11,765 were on time or early. The average highest delay time belongs to Charles M. Schulz-Sonoma County Airport.

Alaska Airlines is less likely to have major delays in comparison to the rest of the industry, probably due to the more regionalized routes and number of available planes at their disposal. This also led to Newark, NJ to San Francisco, CA being the worst route the airline offers and best route the airline offers being Juneau, Alaska to Gustavus, Alaska.

Summary Tables

This summary table describes, for Alaskan Air, the relationship between the origin city and delays, both arrival and departure. Here we can see that New York, NY, has an average total delay of roughly 34 minutes.

Summary Stats for Alaskan Airlines

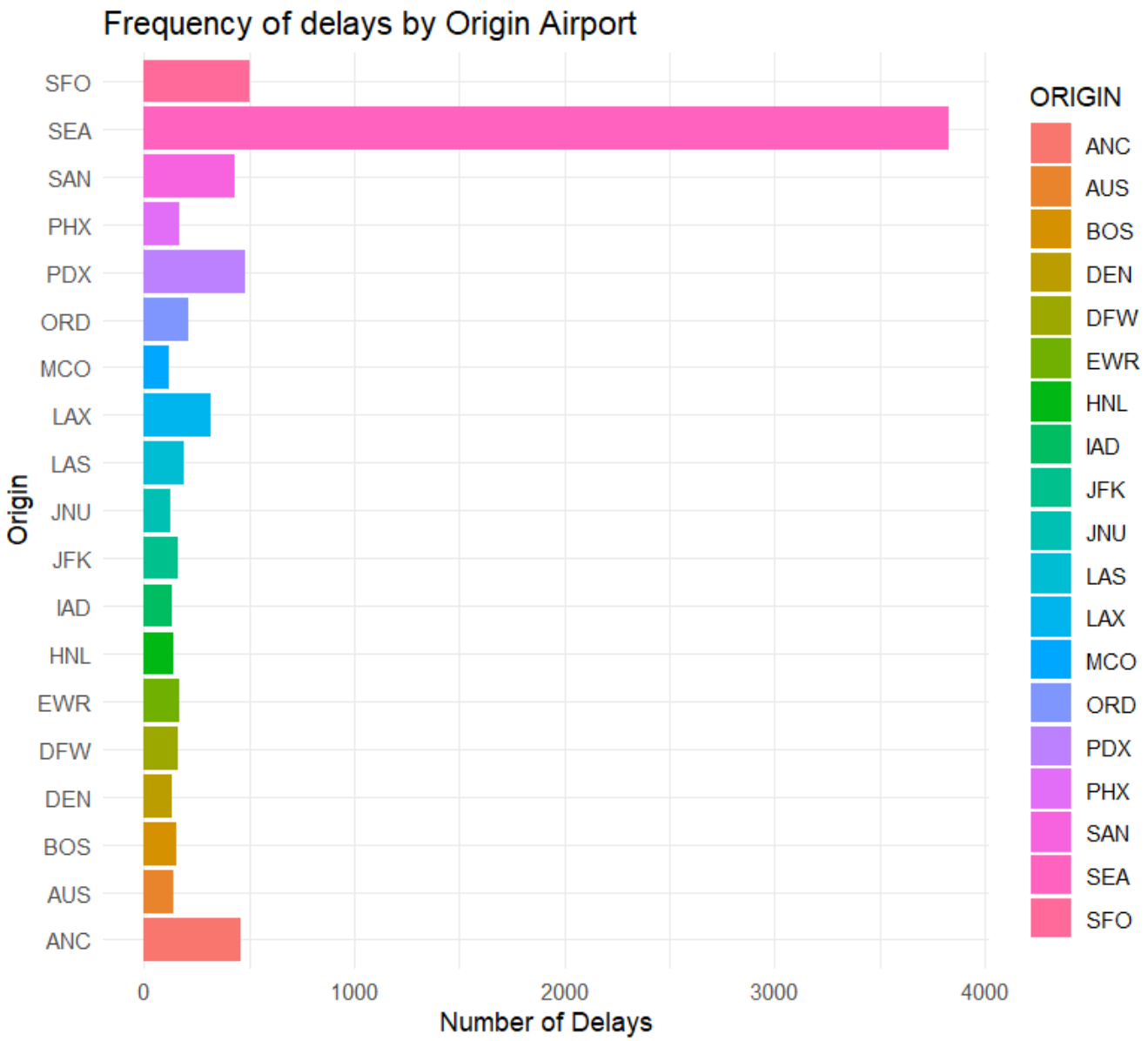
DEST_CITY_NAME	Count	AverageDelayD	AverageDelayA	AverageTotalDelay	DelayDiff
San Francisco, CA	1100	15.085321	16.773897	31.859218	−1.69
Portland, OR	1279	10.198587	11.531496	21.730083	−1.33
Seattle, WA	7162	8.314479	9.385755	17.700233	−1.07
Juneau, AK	462	1.623913	1.725490	3.349403	−0.10
Anchorage, AK	1483	7.248638	6.017747	13.266385	1.23
San Diego, CA	947	7.988335	6.481953	14.470288	1.51
Las Vegas, NV	440	10.127563	8.095672	18.223235	2.03
Phoenix, AZ	360	8.114206	5.821727	13.935933	2.29
Santa Ana, CA	308	4.384868	-1.036424	3.348445	5.42
Chicago, IL	366	12.137363	5.923077	18.060440	6.21
Los Angeles, CA	839	9.694611	3.442446	13.137057	6.25
New York, NY	348	20.447977	13.817391	34.265368	6.63
Honolulu, HI	348	9.593660	2.773913	12.367573	6.82
Washington, DC	390	7.528205	-1.211340	6.316865	8.74
Boston, MA	290	12.993056	2.440972	15.434028	10.55
Newark, NJ	297	19.544218	4.855670	24.399888	14.69

This next summary table describes, also for Alaskan Air, the relationhip between origin and delays.

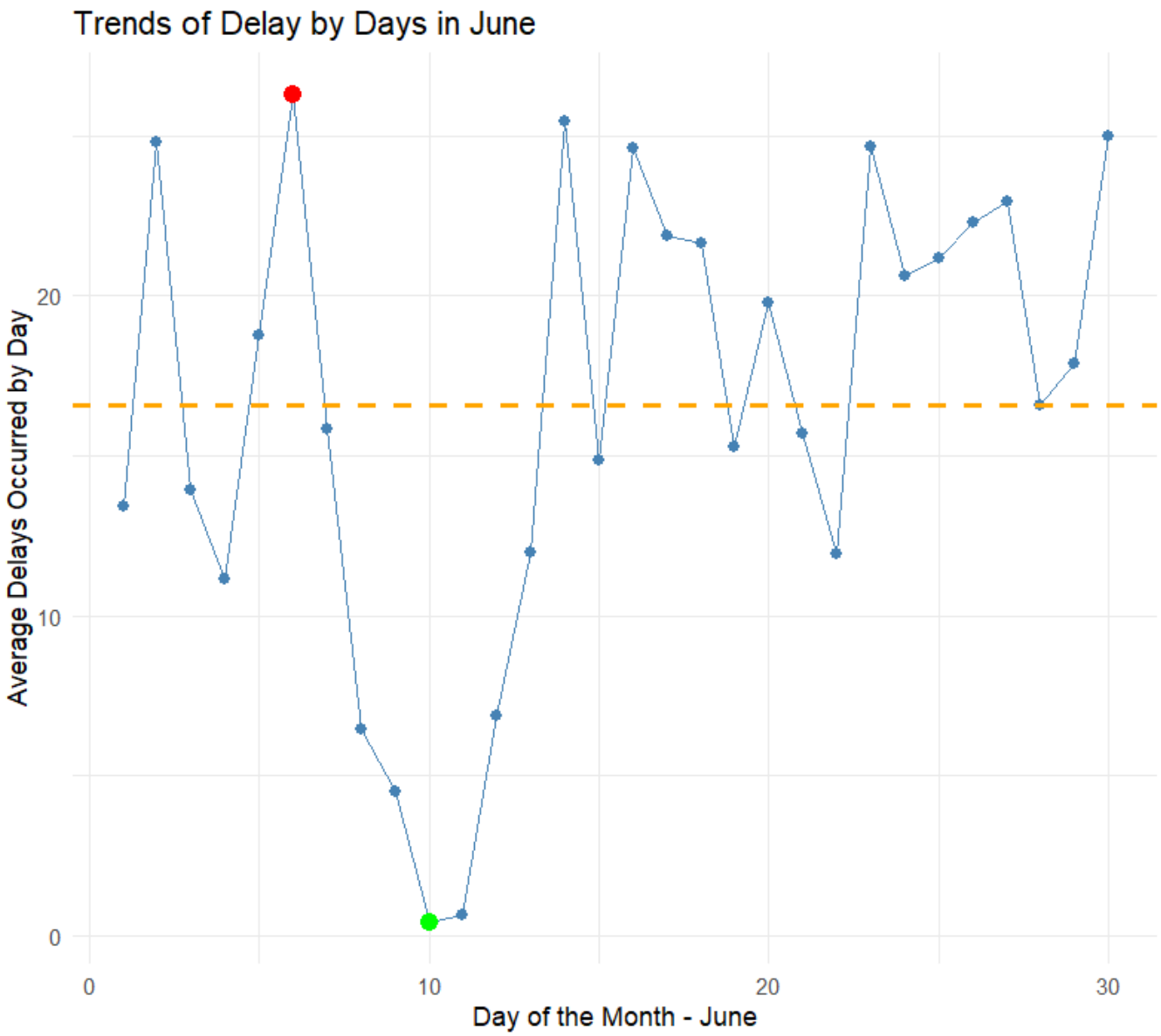
Summary Stats for Alaskan Airlines

ORIGIN_CITY_NAME	Count	AverageDelayD	AverageDelayA	AverageTotalDelay	DelayDiff
Boston, MA	290	12.5929825	18.2147887	30.807771	−5.62
Newark, NJ	297	25.9078498	28.9206897	54.828539	−3.01
Chicago, IL	366	19.8895028	22.5745856	42.464088	−2.69
New York, NY	347	12.8720930	15.1744186	28.046512	−2.30
Phoenix, AZ	362	6.0027701	7.9722992	13.975069	−1.97
San Diego, CA	947	9.7667020	10.4410202	20.207722	−0.67
Washington, DC	390	6.0462725	6.4727273	12.519000	−0.43
Anchorage, AK	1481	0.6976902	0.4597544	1.157445	0.24
Santa Ana, CA	308	2.4750831	1.0099668	3.485050	1.47
Las Vegas, NV	439	8.4654378	6.9124424	15.377880	1.55
Juneau, AK	462	-0.7565217	-2.3485839	-3.105106	1.59
Los Angeles, CA	837	6.8932854	4.3901561	11.283441	2.50
Portland, OR	1279	7.1525157	3.2056738	10.358189	3.95
Seattle, WA	7163	12.0984158	7.6815690	19.779985	4.42
Honolulu, HI	348	12.9505814	7.2238372	20.174419	5.73
San Francisco, CA	1101	13.3649635	7.4358974	20.800861	5.93

Data Visualizations



Above, we can see that for Alaskan Air, SEA has the highest frequency of delays.



Above we can see day 10 in the month of June had the least delays, and day 6 had the most delays.

Business Insights

The airline has minimal delays, and better overall consistency with minimizing delays with service to the west coast, compared to other airlines in the industry.

The service to east coast needs to improve as we can see that flights originating and arriving in the east coast show the highest average delays. Flights arriving and departing from New York and New Jersey had the highest average delays. Here we were only considering the most serviced airports with more than 250 operational flights a month.

They can also investigate potentially forming new partnerships with other brands or airlines to add or service more popular domestic/international flights to help reduce air traffic and delays within its route network. Lastly the airline should continue to find new ways to try to cut down its reduced delay times (16.7min vs 28.46 min industry average) this will help improve customer satisfaction levels and ensure it continues to be one of most highly rated airlines among customers in the U.S. These recommendations will help make Alaska Airlines more competitive and increase revenue without overextending their resources.

Function Report

- Used piping operators to not overwrite the original dataset.
- read_csv(): Reads a CSV file into a tibble (data frame) in R, typically used in the tidyverse environment.
- glimpse(): Provides a compact overview of a dataset, showing the structure and data types of columns.
- filter(): Subsets rows in a data frame based on specified conditions.
- group_by(): Groups data by one or more variables to perform grouped operations, often used with summarise().
- summarise(): Aggregates or summarizes data by applying summary functions (e.g., mean, sum) to columns, usually after grouping.
- mean(): Calculates the arithmetic mean (average) of a numeric vector or column.
- mutate(): Adds or modifies columns in a data frame by applying transformations or calculations.
- arrange(): Orders rows in a data frame based on the values of one or more columns.
- ggplot(): Initializes a ggplot object for creating visualizations, specifying aesthetics and layers for the plot.
- geom_bar(): Adds a bar graph layer to a ggplot visualization, often used for categorical data.
- geom_line(): Adds a line graph layer to a ggplot, typically used for visualizing trends over continuous variables.
- gt(): Creates a table object for displaying data in a formatted, high-quality table using the gt package.