

[◀ Return to Classroom](#)

# Communicate Data Findings

REVIEW

CODE REVIEW

HISTORY

## Meets Specifications

Kudos,

This was a challenging project. Now you have a polished project that you can show case on your Github. Nonetheless, it is also time to take this project to the next level. Please share what you learned with your peers.

At one point you will need to develop a model, train, and then re-train your model from time to time, for this you need to pay attention to the change in data. One way to measure this is called data drift. "Data drift is the sum of data changes — think mobile interactions, sensor logs and web clickstreams — that started life as well-meaning business tweaks or system updates, as CMSWire contributor, Girish Pancha, explains in greater detail here." Below are some interesting links on this:

Some tools like Azure support measuring data drift out of the box ,

- <https://docs.microsoft.com/en-us/azure/machine-learning/how-to-monitor-datasets>
- <https://blog.dominodatalab.com/data-drift-detection-for-image-classifiers/>
- <https://www.welcome.ai/domino-data-lab/post/data-drift-detection-for-image-classifiers>

I invite you to read the resources above, this should get you started.  
Keep up and Good luck

## Code Quality

All code is functional (i.e. no errors are thrown by the code). Warnings are okay, as long as they are not a result of poor coding practices.

Great work so far! You took a lot of variables from the dataset and did an excellent job of systematically exploring it and coming up with some interesting findings.

The project uses functions and loops where possible to reduce repetitive code. Comments and docstrings are used as needed to document code functionality.

Comments were used sparingly, but were useful where they were made. There were enough Markdown cells to keep track of what was happening, though it might have been good to have used comments to introduce each code cell, rather than waiting for after the plot.

## Exploratory Data Analysis

The project appropriately uses univariate, bivariate, and multivariate plots to explore many relationships in the data set. Reasoning is used to justify the flow of the exploration.

You chose to look at a lot of variables! There was some mixing of bivariate plots in the univariate section and a bivariate plot in the multivariate section, but their usages were logical and well-made.

Questions and observations are placed regularly throughout the report, after each plot or set of related plots.

As noted above, you did a good job adding Markdown cells after every few plots. Also mentioned above, it would have been good to include some more text to motivate the plots that follow, rather than just leaving most of the comments after the plots have been made. It'll be good to see more of the logic that you had when going about the data analysis.

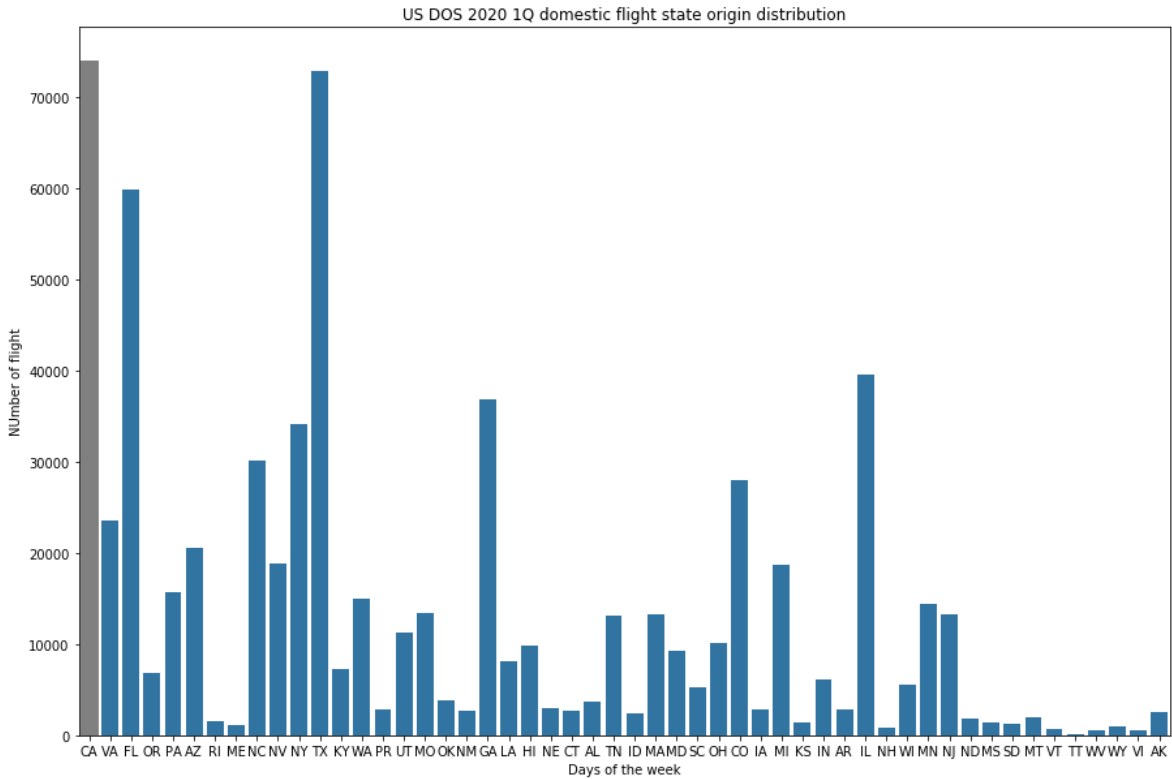
Visualizations made in the project depict the data in an appropriate manner that allows plots to be readily interpreted. This includes choice of appropriate plot type, data encodings, transformations, and labels as needed.

Overall this section is satisfying. However, I will call your attention on some visuals that can be improved.

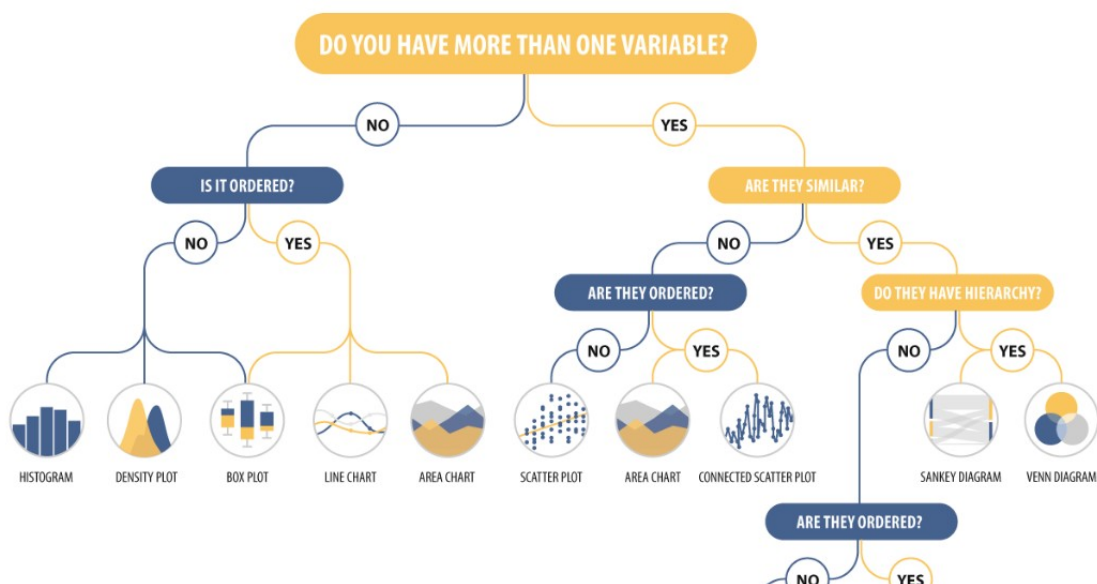
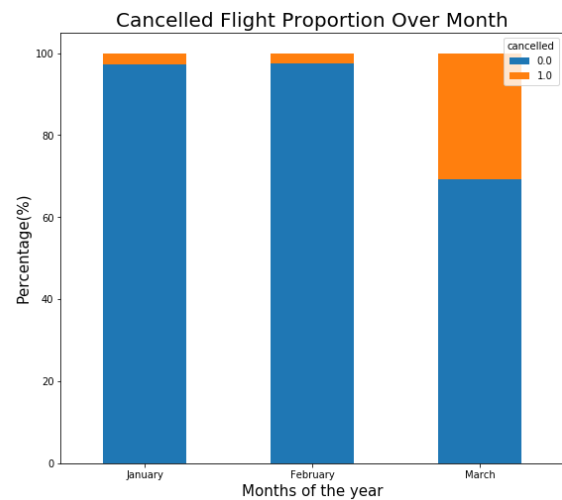
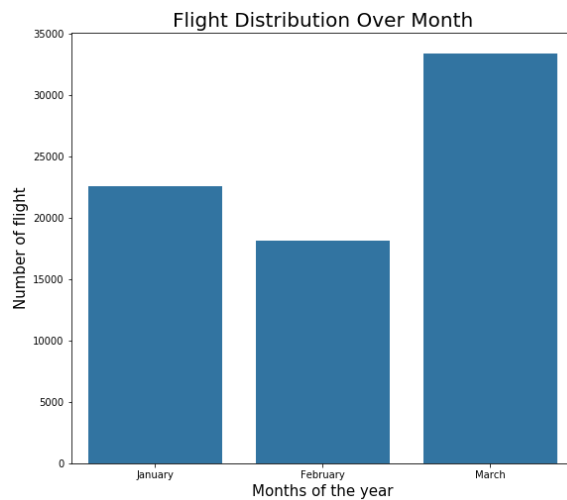
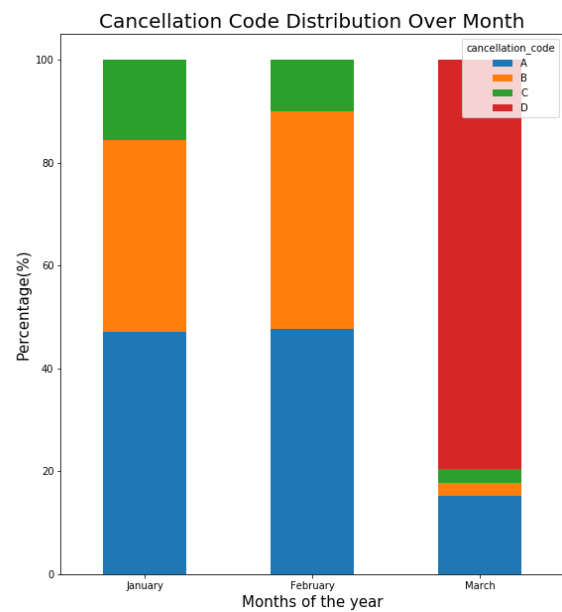
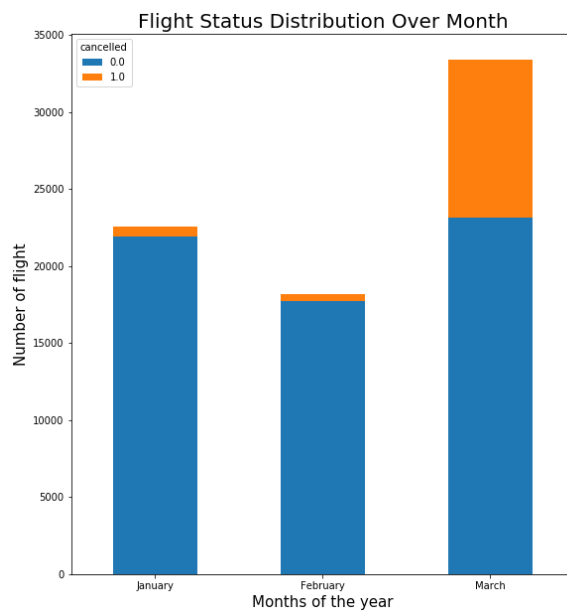
- I like the fact that you highlighted the highest value on your bar chart below, however, it has a wrong x-

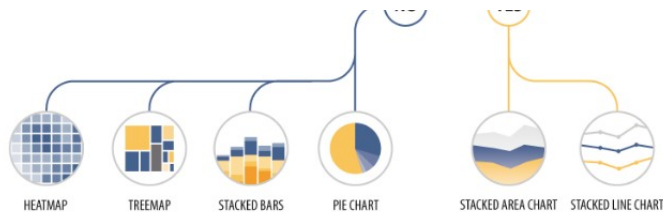
axis title, so this can be confusing to your readers. Moreover, when you have so many values it is important to put some order into it like from highest to lowest or lowest to highest. This helps make

comparison among the different cities faster.



- Using bar chart to represent time series data is incorrect. You always need to use line charts (see the last image below). Line chart is more appropriate because this helps easily visualize the trend.





## Explanatory Data Analysis

A section in the submitted materials includes a summary of main findings that reflects on the steps taken during the data exploration. The section also describes the key insights that are conveyed by the explanatory presentation.

The readme file does a good job of explaining the findings that went into the exploratory analysis and slide deck.

A slideshow is provided, with at least three visualizations used in the presentation to convey key insights. These key insights match those documented in the summary. Each visualization is associated with comments that accurately depict their purpose.

Well done. Your presentation is properly formatted As you can see it is important to convey comments before visuals as this puts the reader into some context to better understand the visual that will follow.

All plots in the presentation have an appropriate title with labeled axes and legends. Labels include units as needed. Plot type, encodings, and transformations are all appropriate.

In addition to comments, it is important to add well-defined labels and titles for your axis and chart, in general, to make sure your readers are focused on understanding the insights you want to share.

[↓ DOWNLOAD PROJECT](#)

[RETURN TO PATH](#)