

1) Import, check and clean the data

2) Filter the data

3) What proportion of taxi journeys start at the airport in total? You can use the following formula to work this out:

$$Proportion_{JFK} = \frac{journeys_{JFK}}{journeys_{all\ locations}}$$

4) Where do people get taxis from in New York? Create a visualization showing the taxi journey starting points.

5) What proportion of taxi journeys start at the airport on each day of the week? Which day of the week shows the highest proportion, and which day shows the lowest? You can use the following formula to work this out:

$$Proportion_{day\ of\ the\ week, JFK} = \frac{journeys_{day\ of\ the\ week, JFK}}{journeys_{day\ of\ the\ week, all\ locations}}$$

6) Create a visualization to show what proportion of journeys take place on each **day of the week**. Do this for journeys from the airport as well as for all locations to see if you can observe any differences. The following formulas represent these two visualizations:

$$Proportion_{day\ of\ the\ week, all\ locations} = \frac{journeys_{day\ of\ the\ week, all\ locations}}{journeys_{all\ days, all\ locations}}$$

$$Proportion_{day\ of\ the\ week, JFK} = \frac{journeys_{day\ of\ the\ week, JFK}}{journeys_{all\ days, JFK}}$$

7) Create a visualization to show what proportion of total journeys are taken each **hour**. Do this for journeys from the airport as well as for all locations to see if you can observe any differences. The following formulas represent these two visualizations:

$$Proportion_{hour, all\ locations} = \frac{journeys_{hour, all\ locations}}{journeys_{all\ hours, all\ locations}}$$

$$Proportion_{day\ of\ the\ week, JFK} = \frac{journeys_{hour, JFK}}{journeys_{all\ hours, JFK}}$$

8) Customize the visualizations

9) Make a recommendation

## Taxi pick-ups in New York



