Excel Instructions

Calculating the duration of trips

1. Add a new column called "duration"

- Right-click the column that you want to insert a new column before. In this case, column `D`.
- Click `Insert Columns`
- Click inside the top cell of the new column
- Type `duration` and press `Enter`

2. Create a formula

- Click in the first empty cell of the `duration` column (`D2`)
- Type `=(`
- Click on the first datetime in the dropoff column (`C2`)
- Type `-`
- Click on the first datetime in the pickup column (`B2`)
- Type `)*1440`

> There are 1440 minutes in a day

- Press `Enter`

3. Change the format of the column

- Right-click the `duration` column (column `D`)
- Click `Number Format...`
- Select `Number`
- Click `OK`

4. Copy the formula down

- Click on the first value in the `duration` column
- Double-click on the small square in the bottom right corner of the cell

Calculating the speed of trips

1. Add a new column called "speed"

- Right-click the column that you want to insert a new column before. In this case, column `E`.
- Click `Insert Columns`
- Click inside the top cell of the new column
- Type `speed` and press `Enter`

2. Create a formula

- Click in the first empty cell of the `speed` column (`E2`)
- Type `=`

- Click on the first value in the `trip_distance` column (`K2`)
- Type `/`
- Click on the first value in the `duration` column (`D2`)
- Type `*60`

> There are 60 minutes in an hour

- Press `Enter`

3. Copy the formula down

- Click on the first value in the `speed` column
- Double-click on the small square in the bottom right corner of the cell

Clean the data

Notice that there are some rows which have a duration or trip distance of zero, and some with a speed of over 100 mph. These are probably not real trips so we can remove them from the dataset to prevent them affecting any later analysis.

1. Sort by duration

- Right-click on the `duration` column name
- Hover over the `Sort` option
- Click `Sort Ascending`

2. Delete rows with zero duration

- Select the first row
- Hold down `Shift`
- Select the last row with a value of zero
- Right click on the selected rows
- Click `Delete Rows`

3. Repeat for trip distance

- Repeat steps 1 and 2 above, this time for the `trip_distance` column

4. Remove high speed trips

- Repeat step 1 above, this time for the `speed` column and using `Sort Descending`
- Repeat step 2 above, this time select the last row with a value greater than `100`

5 Remove values outside July 2021

- Sort by pickup time in ascending order
- Delete rows before July 2021
- Sort by pickup time in descending order
- Delete rows after July 2021

Find hourly means

1. Create a column

- Create a new column after dropoff called `hour`
- Add the formula `=HOUR(B2)`
- Copy the formula down
- Change the `Number Format` of the column to `Number` and change the number of decimal places to `0`

2. Create a pivot table

- Select the `Insert` tab
- Click `PivotTable`
- Click `+ New sheet`

3. Select PivotTable fields

- Drag `hour` to `Rows`
- Drag `duration` to `Values`
- Click the dropdown arrow next to `Sum of duration`
- Click `Value Field Settings`
- Click `Average`
- Click `OK`
- Add average speed and distance to `Values` (as above)

4. Create a plot of speed and distance

- Select the `Insert` tab
- Click the Line chart icon 
- Double-click on the chart to adjust settings