

DS 201X – homework 2

Revisit the HW 1, answer the following questions

Instruction:

Import the data from the **Lemonade2016-2.csv** to Pandas DataFrame

Cleanse the Data

1. Find and remove any duplicate rows in the data.
2. Find and resolve any missing values in the data, either by interpolating missing values that are in an obvious sequence, or by entering the average value for the column (rounded to the nearest whole number) into any cells where a value is missing.

Add Derived Columns

1. Add a column named **Sales** to the table, in which the total sales of lemon and orange is calculated.
2. Add a column named **Revenue** to the table, in which the sales revenue is calculated by multiplying the total sales and price.

Create Charts

1. Create a scatter-plot chart that shows **Leaflets** on the X axis and **Sales** on the Y axis.
2. Create a histogram that shows **Revenue** distributed into 10 bins, and note whether the distribution of this data is normal, left skewed, or right skewed.
3. Create Box and Whisker plot for two **Location** (one box for Park, one for Beach) showing **Total Sales** values
4. Create Box and Whisker plot for two **Location** (one box for Park, one for Beach) showing **Total Sales** values. For only the sales of the day that Temperature was at least 75 F

Show your chart in class.

Also submit your Jupyter Notebook file (.ipynb file) on Canvas.