Data cleaning documentation

Data analyst: Abraham Cedeño

Date: 6th October 2022

Version #: 2

Steps:

Date-time: 6/10/22-11:40am (GMT-5)

1. I made a copy of all the xlxs files and named them as version 2 (v2).

Date-time: 6/10/22-12::03pm (GMT-5)

1. I opened the files and double-checked if there were any duplicates (none were found)

Date-time: 6/10/22-12:11pm (GMT-5)

1. I’ve checked for blank spaces in all the data with the COUNTIF and the LEN functions.

Date-time: 6/10/22-12:26pm (GMT-5)

1. I’ve checked for blank spaces in all the data with the COUNTIF and the LEN functions.

Date-time: 6/10/22-1:15pm (GMT-5)

1. I’ve used the int function to separate date and time in different columns.

Date-time: 7/10/22-12:22pm (GMT-5)

1. I’ve named each column subsequently: started date, started time, end date, end time.

Date-time: 7/10/22-12:34pm (GMT-5)

1. I’ve used the function weekday to create a column with the day of the week of each trip.

Date-time: 7/10/22-12:49pm (GMT-5)

1. I’ve subtracted the end time and started time columns to create a new column which I called ride length, then I sum +1 to this result to avoid problems with negative results.

Date-time: 7/10/22-12:55pm (GMT-5)

1. I saved all data as csv and xlsx files.

Date-time: 7/10/22-1:00pm (GMT-5)

1. I protected worksheets so that no one can access them without the password. Also, I protected the whole excel file so that only people who know the password can open it.