**Statistical and Hypothetical Questions**

* **Outcome of your EDA**

Road traffic trends during 2020 have been affected by the coronavirus (COVID-19) pandemic in the UK. 280.5 billion vehicle miles were driven on Great Britain’s roads in 2020, a decrease of 21.3% compared to the previous year. Traffic statistics are mostly presented in units of vehicle miles, which combines the number of vehicles on the road and how far they drive. As shown in the analysis, vehicle miles travelled in Great Britain have had year-on-year growth in each year between 2010 and 2019. However, the sharp decrease in 2020 has resulted in traffic estimates that are lower than the 2010 levels. Therefore, to say traffic has fallen over the last decade would misconstrue as the overall decrease is entirely due to the decline in traffic levels observed in the 2020 estimates.

Thirteen variables are available in this region traffic by vehicle type dataset, which I will explore five of those variables on this analysis.

Below is the list of 5 variables belong description in this dataset which I’ve used during my analysis.

* Pedal Cycle is a float cyclist on public highways and the paths next to them. This is the lowest level of cycling on the public highway since the 1990s.
* Two Wheeled Motor Vehicles is a floating-point value during the pandemic.
* Cars and Taxis is the integer type decreased in traffic level during the pandemic.
* Buses and Coaches is a float type that were most heavily impacted during COVID 19 pandemic.
* All motor Vehicles is an integer for the outcome of the pandemic time which is lowest annual estimate of car traffic in the last 29 years.
* **What do you feel was missed during the analysis?**

Road accident and safety statistics during the COVID 19 pandemic was a first variable I missed on this analysis. I was struggling to find the related data set to add on, but unfortunately could not find it.

National Atmospheric Emissions Inventory (NAEI) was a second variable I missed on this analysis. Road traffic statistics are used to produce the National Atmospheric Emissions Inventory (NAEI), a legal requirement for EU Air Quality Directives, and for the UN Framework Convention on Climate Change. I wished I could find the NAEI data set of quality of air pollution during the pandemic to compare to previous years.

* **Were there any variables you felt could have helped in the analysis?**

Probability Mass Function (PMF) could help me to represent a distribution on selected variables. I wish I could use PMF for all variables to get more sense of my analyses.

* **Were there any assumptions made you felt were incorrect?**

Based on my knowledge of analysis and very clean data set, I think I haven’t faced any incorrect assumptions.

* **What challenges did you face, what did you not fully understand?**

The biggest challenges l had faced was finding appropriate code, which I’m new in this field (I have mechanical engineer background with zero experience coding) and I think I’m far behind the class. Importing the thinkstats2 package to run my code was kind of challengeable. Hypothesis testing is one of the most difficult topics in an introductory statistic for me. I was really struggling to understand what P- Values are.