Data Science Mini Project

There are 2 questions in this mini project in which you are supposed to submit a solution. Your solution must meet the following criteria :

1. The code must be available in a Github Repo and shared with us.
2. The code must contain documentation on how to run and where to start
3. Make sure to include a requirements file if you use Python.
4. We strongly encourage using Boilerplates and templates like Cookiecutter and etc.

# 

# Notes

* **You have until end of this week (Friday Farvardin 22th)**
* **Time is an important factor. If you submit the answers sooner, you'll have an advantage.**
* A good Data Scientist will finish this exam in 2 hours.
* We use Python but feel free to do it in whatever language you prefer.
* Please make a new commit for each step.
* When finished, please push to GitHub and send us a link.

***Send your final answers as a Github Link to : p.esfandiari@uzer.one,***

***a.ebrahimi@uzer.one***

# Question 1 :

Consider a deck of cards with **N** different playing cards equally distributed among **M** suits. For example, a standard deck of card has suits with cards for each suite. You draw all the cards without putting any back. After the first card, each time you pick a card, you compare it to the previous card. If the suits match, you get a point. Otherwise, you get no point. Please answer the following questions about the value of P (amount of points you get at the end of this process) :

1. What is the mean of when and ?
2. What is the Standard Deviation of for and ?
3. What is the mean of when and ?
4. What is the Standard Deviation of for and ?
5. What is the conditional probability that given that it's when and ?
6. What is the conditional probability that given that it's when and ?

# Question 2 :

Consider the following Datasets available in the data folder. In order to understand the data, the link below is provided to understand the data model:

<https://github.com/stanford-policylab/opp/blob/master/data_readme.md>

Given the two datasets above, please provide the answer to the following queries with the code:

1. The proportion of traffic stops in MT involving male drivers
2. Factor increase in a traffic stop arrest likelihood in MT from OOS plates
   1. Chi-Squared traffic stop arrest test statistic
3. The proportion of traffic stops in MT involving speeding violations
4. Factor increase in traffic stop DUI likelihood in MT over VT:
5. The average manufacture year of vehicles stopped in MT in 2020
   1. P-value of linear regression:
6. The difference in the total number of stops that occurred between min and max hours in both MT and VT
7. The area in sq. km of the largest county in MT