The problems to practice basic probability

These problems can be solved either manually or with code, as you wish. In any case, an analytic solution (not a numeric approximation) is required, unless otherwise specified.

The results can be presented as a scan of your handwritten solution, or in a digital format such as pdf, doc or ipynb. In any case, the results should be submitted to Google Classroom.

Stock market (20 points). The price of the ABCD stock is $256 today. Every day the stock price goes either up by 25%, or down by 20%, with equal probabilities, and all days are independent from each other. A call option allows buying the stock for $300 exactly in 4 days. How much is this option worth?

Hitchhiker (20 points). The probability of at least one car passing a certain road intersection in a 20-minute window is 0.9. What is the probability of at least one car passing the intersection in a 5-minute window, assuming that cars pass this intersection with a constant rate?

Credit scoring (20 points). A new client applies to a bank for a consumer loan. She has a recent application in another bank, and she has repaid a mortgage loan two years ago. We want to estimate the probability that she is a "bad" client (she will not repay the loan). We believe that recent applications and repaid loans are independent *conditionally* on the client being bad/good (but not independent in general) – i.e. we make the naïve Bayes assumption. We know that:

- average share of bad clients is 10%;
- 80% of bad clients have recent applications in other banks, but only 30% of good clients have such applications;
- 10% of good clients have a mortgage repaid, but only 5% of bad clients have it.

So what is the probability that she is bad?

The programming problem (40 points)

- Look at the code that produced descriptive statistics and plots for the lecture: https://colab.research.google.com/drive/1TnggXJe_7GviZmr13D_xjnkIxfU6uxCz?usp=sharing
- Find a dataset of your choice with at least two categorical and at least two numeric meaningful variables. Ideally, one of these variables should be a potential target for prediction. If you do not have such a dataset, choose one on Kaggle or use the dataset credit.csv from https://github.com/stedy/Machine-Learning-with-R-datasets.
- Create a Jupyter notebook that explores this dataset using descriptive statistics and plots, and formulate the key insights you found in the data.
    - Load the dataset, show a small sample of the data and explain its nature
    - Select a few variables of interest and formulate meaningful questions or hypotheses about them
    - Use descriptive statistics and visualization to answer these questions
    - You don't have to submit the dataset itself, but you have to submit the code with which you analyze it, the output of this code, and your comments.