AKATA GOAVANA

Data Analysis Training

Jul 10, 2024

Create a public Git repository (GitLab or GitHub)

- Repository Name: data-analysis-train-{name}
- Follow the company's coding standards
    - Naming of variables, functions, classes, etc.
    - Add comments to clarify the code.
- Follow the company's commit standards
    - Commit message structure
    - One commit per feature (Try to separate commits)
- Add the processes to follow for testing your work in the README (DOCS)
- Send the link to your repository as soon as you are finished; we also want to assess your speed.

Data

Download the files train.csv.zip and bids.csv.zip from https://www.kaggle.com/c/facebook-recruiting-iv-human-or-bot/data

Problem definition

Facebook is interested in understanding the difference between bidding behaviour of humans and of robots.

Task 1: analysis

Using the python library pandas, calculate the following:

1. For each auction, the number of unique ip addresses

2. For each merchandise type, the country which sent the maximum number of bids (in the csv, include also the number of bids sent by that country)

3. The data contains three separate time periods. Calculate average number of unique bidders per auction in each of the three separate time periods

4. For each country, the fraction of times a bid from that country was identified as coming from a robot

Output for sending to Git:

• 4 csv files (one per question, called question_1.csv etc.)

• code used to produce the output (.ipynb)

Task 2: analysis and visualization

Using any programming language and libraries you want, perform exploratory data analysis on the data. You choose the variables you want to look at, the new features you want to construct, etc.

Output for sending to Git:

• a series of graphics with short explanatory texts about the interesting conclusions you draw from each graphic

• code used to produce the graphics (e.g. in an ipynb notebook)