

Intorduction

This assignment is designed to help you practice your data parsing and feature engineering skills. You will be given a dataset called `data.csv` that contains three columns: `id`, `date_start`, and `contracts`. The `contracts` column is a JSON string that contains information about the contracts that a customer has signed up for. Your task is to parse this JSON string and calculate a set of features from it.

Data

The `data.csv` file contains the following columns:

- `id`: The unique identifier for the row.
- `date_start`: The start date of the application.
- `contracts`: A JSON string that contains information about the contracts that a customer has signed up for. You must analyze JSON data format yourself.

Deliverables

- Code/script files: Provide the script or code (e.g., Python script, Jupyter notebook) that performs the necessary parsing and feature calculation.
- Output dataset: Share the resulting dataset in csv file that includes the calculated features as columns. Please name it as `contract_features.csv`.

Additional Notes

- You can modify the assignment to calculate more features.
- You can only use python for this assignment.
- Use the pandas library to load and manipulate the data.
- Use the json library to parse the JSON strings.
- You can find the definitions of the features in the `features.xlsx` file.

Let us know if you have any other questions.

Submission

You must share a link to a version-controlled repository (e.g., GitHub).

We look forward to reviewing your work and discussing it with you. Good luck!