ta-data-lis / Project-Week-3-Data-Thieves

Code Issues Pull requests 1 Actions Projects Wiki Secu

master Go to file Add file Code

carlarsmendes committe... 4 commits 1 branch 0 tags

your-project	first commit	11 months ago
Readme.md	changes on readme ideas	11 months ago

About

No description, website, or topics provided.

Readme

Releases

No releases published

Readme.md

Welcome to the Data Thieves Project!

In this project you will get all the data yourself! 😱 Get ready to present your Project on Friday :)

Content

Project Description

In this project, you will choose a topic and find all the relevant data yourself from Kaggle. Afterwards, you should enrich it by connecting to an API, find a dataset or scrape data from the web. You then must organize, clean and analyse the data you find and present your findings in a presentation (you may use plots!).

Project Goals

- Learn how to develop an interesting question and find the data to answer it.
- Learn how to obtain data from different sources, including APIs, open source datasets and/or scrape data from the web.
- Build a database from the data you find for the whole team to use.
- Explain more complex arguments with plots.

Requirements

- You must plan your project. That is why creating a Kanban or Trello Board is mandatory. You have a template for Trello here.
- You **CAN'T CODE** until you project is planned.
- Create a *.gitignore* file and include it in your repository.
- Your project must include data from (at least) 2 different data sources (APIs & web, dataset & APIs, ...)

Deliverables

You are required to turn in the following:

1. Link to your profile on Kaggle.
2. Link to your GitHub Repository and README.
3. Documentation as talked about in class.
4. Access information to your database with a description of each table and how they relate.
5. Links to the data you are using (sources) and your organization (trello).
6. Slides for your presentation.

Mentoring

Either a TA or the Lead Teacher will be your mentor! Your mentor will:

- Follow your project in general, will be the second person that knows more about the project, after you.
- Check if you are following the tasks, your blockers, etc
- Help/support you in specific questions.

Your mentor is NOT meant to:

- Know everything.
- Be your manager. You have to be the responsible person to do the tasks!

Schedule

Tuesday & Wednesday

- Look for an interesting topic and make some hypothesis or think about some questions to answer about it.
- Investigate which data sources are available for that topic.
- Reach some best practices agreements as a team.
- Plan your project and organize. Think about some risks you can expect.
- Start working on your database.

Thursday

- Start working on your analysis and plots. Think about the plots you want to create and the structure of your presentation.
- Finish your analysis.
- Start working on your presentation.

Friday

- Adjust your presentation.
- Presentation!!

README File structure

The README will be your paper and it is meant to have all the (analysis) information about your project.

The structure should be:

1. Title of the project
2. Introduction to your project.
3. Data you are using (and comments, main challenges, strengths & weaknesses, etc...)
4. Questions you want to answer (maybe divided by different topics). Each question should include a conclusion written in a markdown cell.
5. Conclusions after your analysis.
6. Further questions.

Presentation

You will have 10 minutes to present your project. The below are some ideas for slides you could include in your presentation; those marked with an **(M)** are mandatory!

- **(M)** Title of the project
- **(M)** Your topic. Why did you choose it?
- **(M)** Presentation of the team

- Main challenges & strengths
- **(M)** Team. Did you follow your workflow plan? Did you add something after starting the project? Did you follow your best practices agreements? Did you think about the risk management?
- About your data: useful sources, incomplete data, data that would have been great to have, etc.
- Data cleaning: how and why you cleaned your data the way you did.
- **(M)** Main insights: one slide per insight!
- Questions you couldn't answer.
- Something funny that happened during the project.
- Things you learned during this project.
- If you could start from scratch, what would you do differently?

Tips & Tricks

- First, choose your topic and look for sources available.
- Before you start coding and integrate more data, propose some interesting questions you could answer with the data you have.

Resources

Lists

AnyAPI
Top 50 Most Popular APIs on RapidAPI
18 Fun APIs For Your Next Project

Some Ideas

WeatherBit
Strava
GitHub
Twitter
LastFM
Spotify
NYTimes
News
Reddit
Medium
Twitch
IGDB
OMDB
GIPHY
StackExchange
YouTube
TheSportsDB
NBA API

Paper Examples

Data Analysis with Python
The Best Mario Kart Character According To Data Science