

CSCE 5320 Scientific Data Visualization

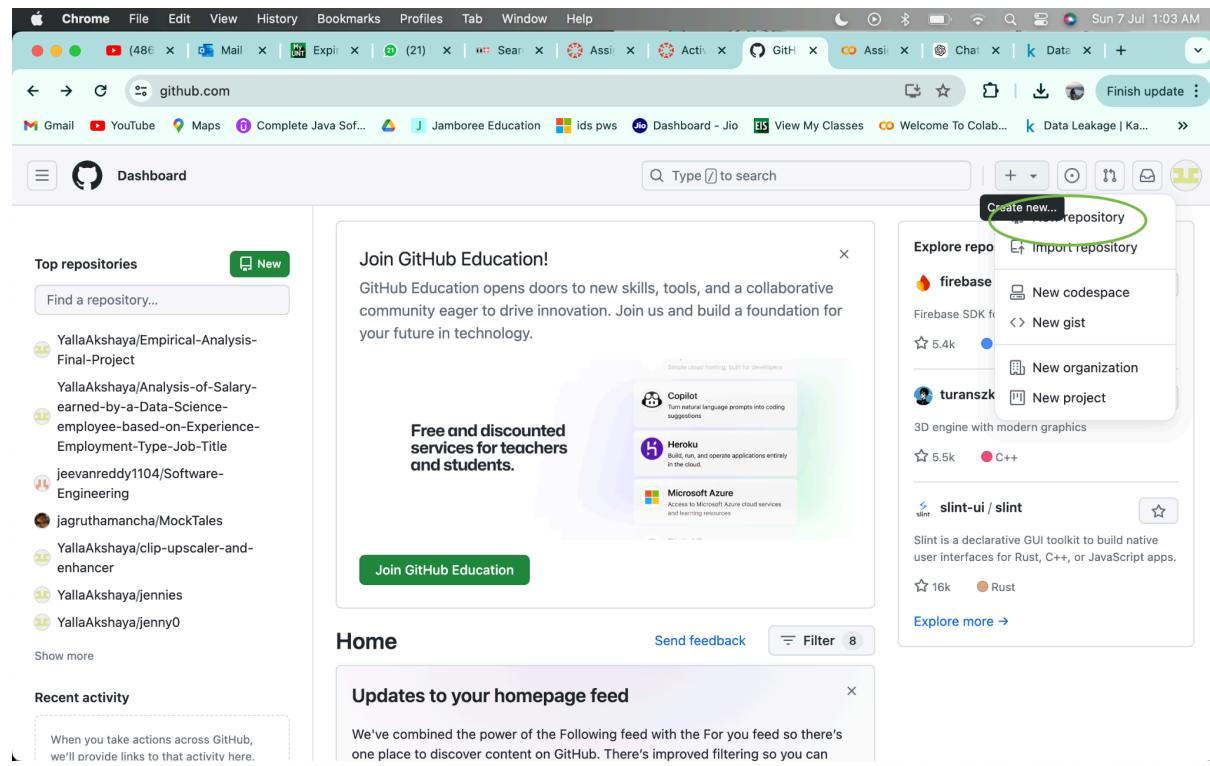
Assignment 2

Question 1 :

Here, we are creating a pie chart to visualise a dataset using d3.js dataset and vizhub.

Firstly, let us create a link of our dataset using github, so that it can be accessed easily to make visualisations.

Open github.com, and click on the + icon on top right. Then, click on the create new repository from the drop down menu.



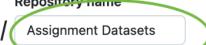
Then, I have taken the repository name as 'Assignments datasets', chose it as public so that it can be accessed from anywhere and by everyone, select the 'add readme file' and click on done.

Chrome File Edit View History Bookmarks Profiles Tab Window Help Sun 7 Jul 1:04 AM

github.com/new

A repository contains all project files, including the revision history. Already have a project repository elsewhere? [Import a repository.](#)

Required fields are marked with an asterisk (*).

Owner * YallaAkshaya **Repository name *** Assignment Datasets 

Your new repository will be created as Assignment-Datasets. The repository name can only contain ASCII letters, digits, and the characters ., -, and _.

Great repository names are short and memorable. Need inspiration? How about super-duper-fishstick ?

Description (optional)

Public Anyone on the internet can see this repository. You choose who can commit. **Private** You choose who can see and commit to this repository.

Initialize this repository with: Add a README file This is where you can write a long description for your project. [Learn more about READMEs.](#)

Add .gitignore .gitignore template: None Choose which files not to track from a list of templates. [Learn more about ignoring files.](#)

Choose a license License: None

Then, a repository is created. Now, click on upload files which can be seen when 'add files' button is clicked.

Chrome File Edit View History Bookmarks Profiles Tab Window Help Tue 9 Jul 12:30 PM

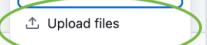
github.com/YallaAkshaya/Assignment-Datasets

YallaAkshaya / Assignment-Datasets Type ⌘ to search

Code Issues Pull requests Actions Projects Wiki Security Insights Settings

Assignment-Datasets Public

main 1 Branch 0 Tags Go to file Add file Code

+ Create new file 

YallaAkshaya Initial commit

README.md Initial commit

README

About

No description, website, or topics provided.

Readme Activity 0 stars 1 watching 0 forks

Releases

No releases published [Create a new release](#)

Packages

No packages published [Publish your first package](#)

<https://github.com/YallaAkshaya/Assignment-Datasets/new/main>

Click on choose your files option as shown below. And upload it from your computer as I did which is the ‘salary predictions of data professions’, which I downloaded from kaggle.

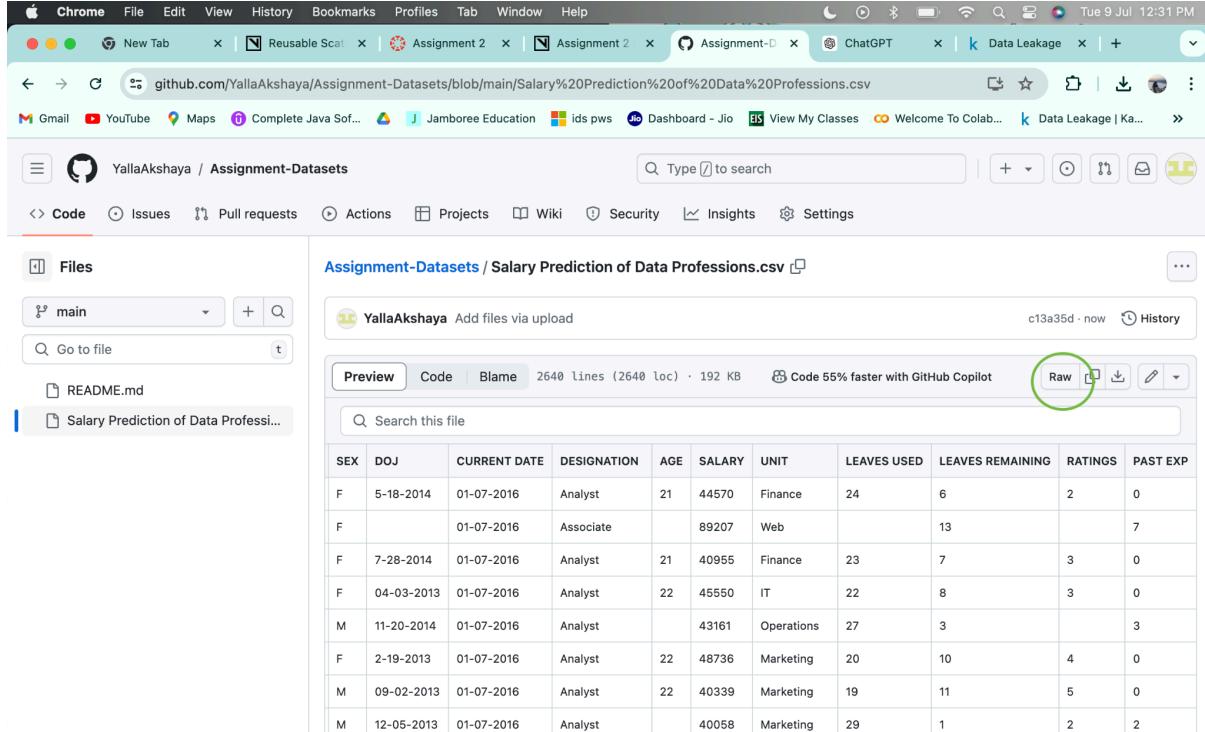
<https://www.kaggle.com/code/mariabassam/salary-prediction-of-data-professions/notebook>

Then, click on ‘commit changes’

The screenshot shows a GitHub commit dialog. At the top, there is a large input field with a 'Drop files here to add them to your repository' placeholder and a 'Or choose your files' link. Below this is a 'Commit changes' section with fields for adding files via upload and an optional extended description. There are two radio button options for committing: one to the 'main' branch and another to create a new branch. The 'choose your files' link is circled in green. In the second part of the dialog, a file named 'Salary Prediction of Data Professions.csv' is selected and also circled in green. At the bottom, there are 'Commit changes' and 'Cancel' buttons.

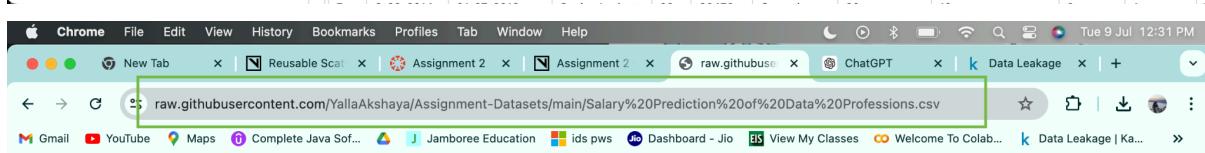
Then, open the csv file which we have uploaded and click on 'raw'.

With this, a link will be created, which redirects directly to the webpage, where the content of the dataset is present as a csv file.



The screenshot shows a GitHub repository page for 'Assignment-Datasets'. On the left, there's a sidebar with 'Files' containing 'main', 'README.md', and 'Salary Prediction of Data Professions..csv'. The main area displays the contents of the CSV file. At the top of the CSV preview, there's a toolbar with 'Preview', 'Code', 'Blame', 'Raw' (which is circled in green), and other options. Below the toolbar is a search bar. The CSV data starts with:

| SEX | DOJ | CURRENT DATE | DESIGNATION | AGE | SALARY | UNIT | LEAVES USED | LEAVES REMAINING | RATINGS | PAST EXP |
|-----|------------|--------------|-------------|-----|--------|------------|-------------|------------------|---------|----------|
| F | 5-18-2014 | 01-07-2016 | Analyst | 21 | 44570 | Finance | 24 | 6 | 2 | 0 |
| F | | 01-07-2016 | Associate | | 89207 | Web | | 13 | | 7 |
| F | 7-28-2014 | 01-07-2016 | Analyst | 21 | 40955 | Finance | 23 | 7 | 3 | 0 |
| F | 04-03-2013 | 01-07-2016 | Analyst | 22 | 45550 | IT | 22 | 8 | 3 | 0 |
| M | 11-20-2014 | 01-07-2016 | Analyst | | 43161 | Operations | 27 | 3 | | 3 |
| F | 2-19-2013 | 01-07-2016 | Analyst | 22 | 48736 | Marketing | 20 | 10 | 4 | 0 |
| M | 09-02-2013 | 01-07-2016 | Analyst | 22 | 40339 | Marketing | 19 | 11 | 5 | 0 |
| M | 12-05-2013 | 01-07-2016 | Analyst | | 40058 | Marketing | 29 | 1 | 2 | 2 |



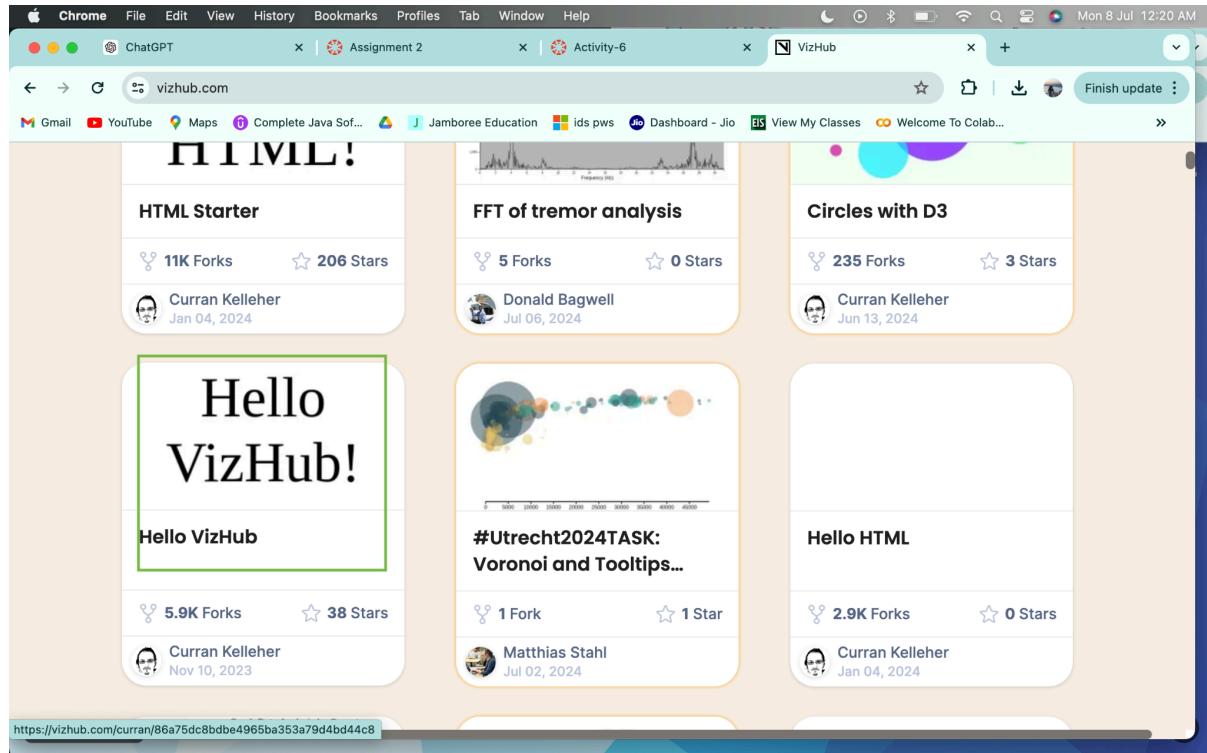
The screenshot shows a browser window with the URL 'raw.githubusercontent.com/YallaAkshaya/Assignment-Datasets/main/Salary%20Prediction%20of%20Data%20Professions.csv'. The page displays the same CSV data as the GitHub preview, starting with:

| SEX | DOJ | CURRENT DATE | DESIGNATION | AGE | SALARY | UNIT | LEAVES USED | LEAVES REMAINING | RATINGS | PAST EXP |
|-----|------------|--------------|-------------|-----|--------|------------|-------------|------------------|---------|----------|
| F | 5-18-2014 | 01-07-2016 | Analyst | 21 | 44570 | Finance | 24 | 6 | 2 | 0 |
| F | | 01-07-2016 | Associate | | 89207 | Web | | 13 | | 7 |
| F | 7-28-2014 | 01-07-2016 | Analyst | 21 | 40955 | Finance | 23 | 7 | 3 | 0 |
| F | 04-03-2013 | 01-07-2016 | Analyst | 22 | 45550 | IT | 22 | 8 | 3 | 0 |
| M | 11-20-2014 | 01-07-2016 | Analyst | | 43161 | Operations | 27 | 3 | | 3 |
| F | 2-19-2013 | 01-07-2016 | Analyst | 22 | 48736 | Marketing | 20 | 10 | 4 | 0 |
| M | 09-02-2013 | 01-07-2016 | Analyst | 22 | 40339 | Marketing | 19 | 11 | 5 | 0 |
| M | 12-05-2013 | 01-07-2016 | Analyst | | 40058 | Marketing | 29 | 1 | 2 | 2 |

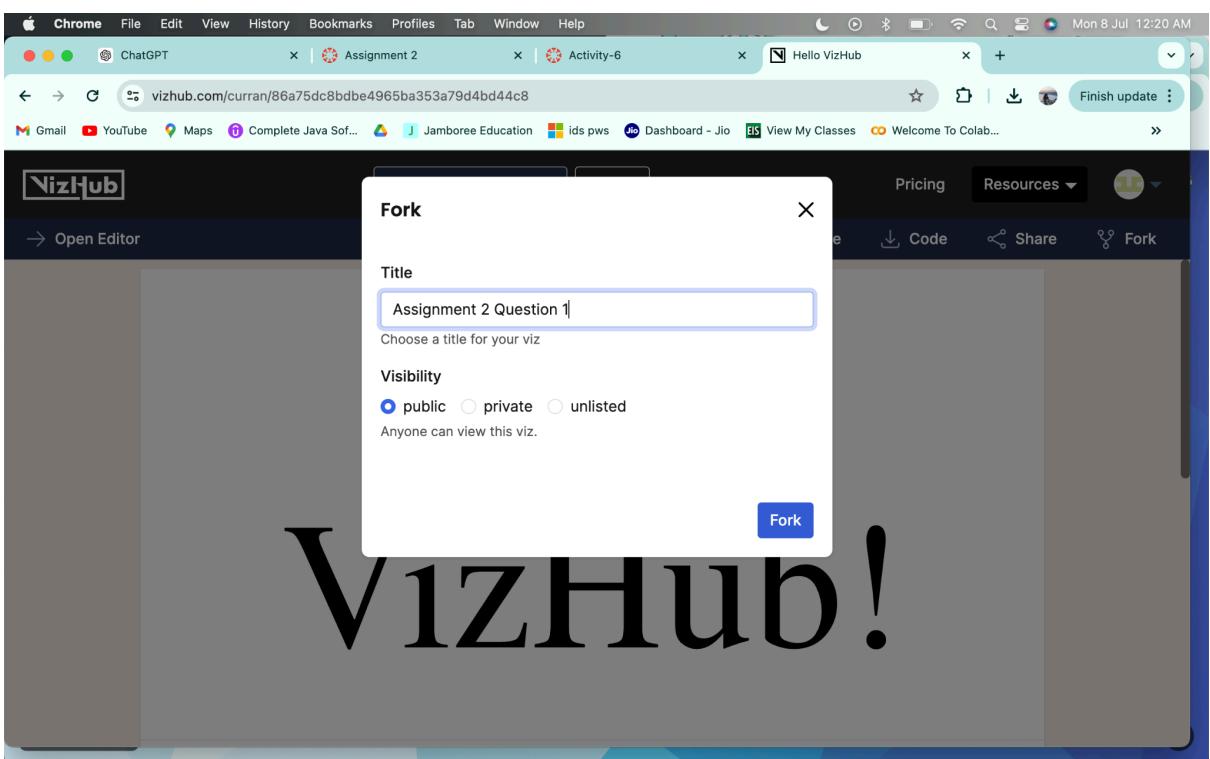
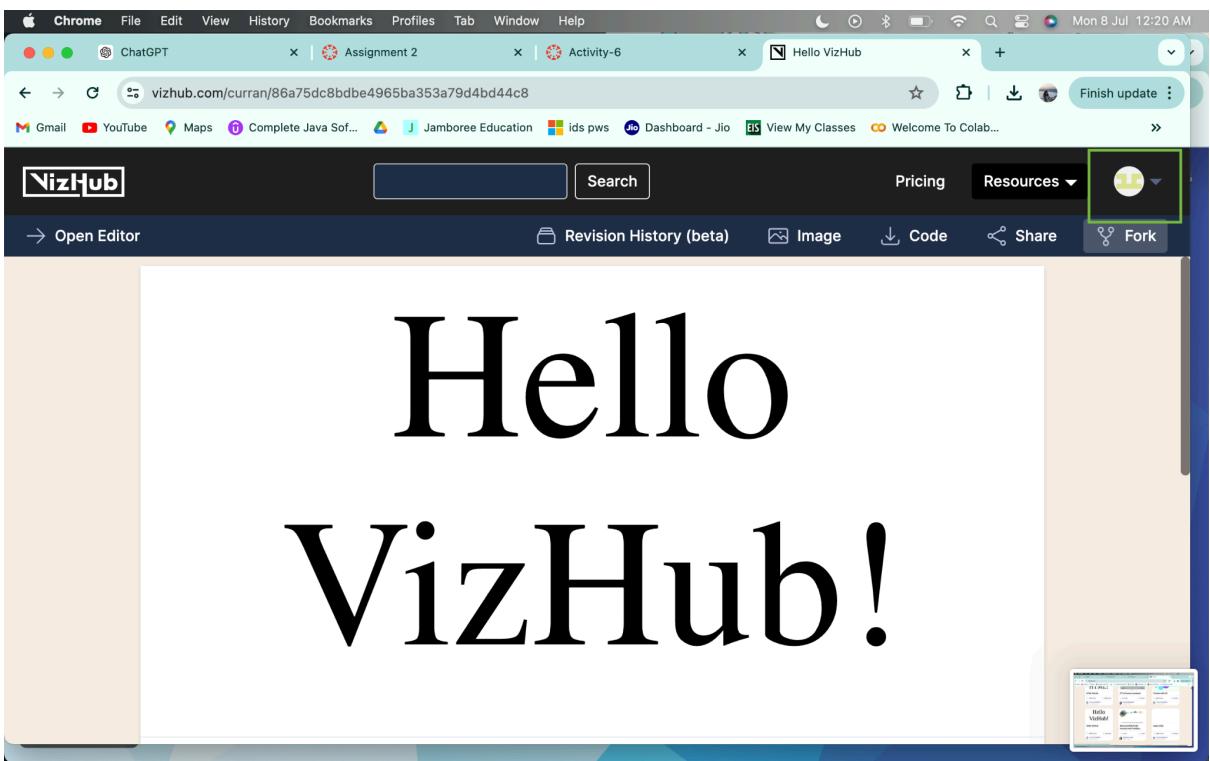
Github link:

<https://raw.githubusercontent.com/YallaAkshaya/Assignment-Datasets/main/Salary%20Prediction%20of%20Data%20Professions.csv>

Now, click on the <https://vizhub.com/>, login to it by connecting it with your github account and select 'hello vizhub!' one as shown in the screenshot below.



Now, click on the fork symbol, so that you can save the project and make changes to it. Give the title and choose the visibility. Choose it to the public as it needs to be accessed by everyone who has the link. And then, click on the fork button.



Write the appropriate code to visualise the dataset using pie charts in js and html files. Making sure, that the code should work for the url of the dataset we have created using github

Chrome File Edit View History Bookmarks Profiles Tab Window Help

vizhub.com/YallaAkshaya/15e3ea368a5d42c79ac8d98567bb5a9a?edit=files&file=index.js

index.html index.js README.md

index.js

```

1 const url =
2   'https://raw.githubusercontent.com/YallaAkshaya/Assignment-Datasets/main/'
3
4 // Fetch the data from GitHub
5 d3.csv(url).then(function (data) {
6   // Process data to count each unit
7   const unitCount = d3.rollup(
8     data,
9     (v) => v.length,
10    (d) => d.UNIT,
11  );
12  const processedData = Array.from(
13    unitCount,
14    ([unit, count]) => ({ unit, count })
15  );
16
17  // Calculate total count for percentage calculation
18  const totalCount = processedData.reduce(
19    (acc, curr) => acc + curr.count,
20    0,
21  );
22
23  // Dimensions and radius
24  const width = 900,
      height = 600

```

Connected

Revision History (beta) Image Code Share Settings Delete Fork

Distribution of Employees Across Different Units

Assignment 2 Question 1

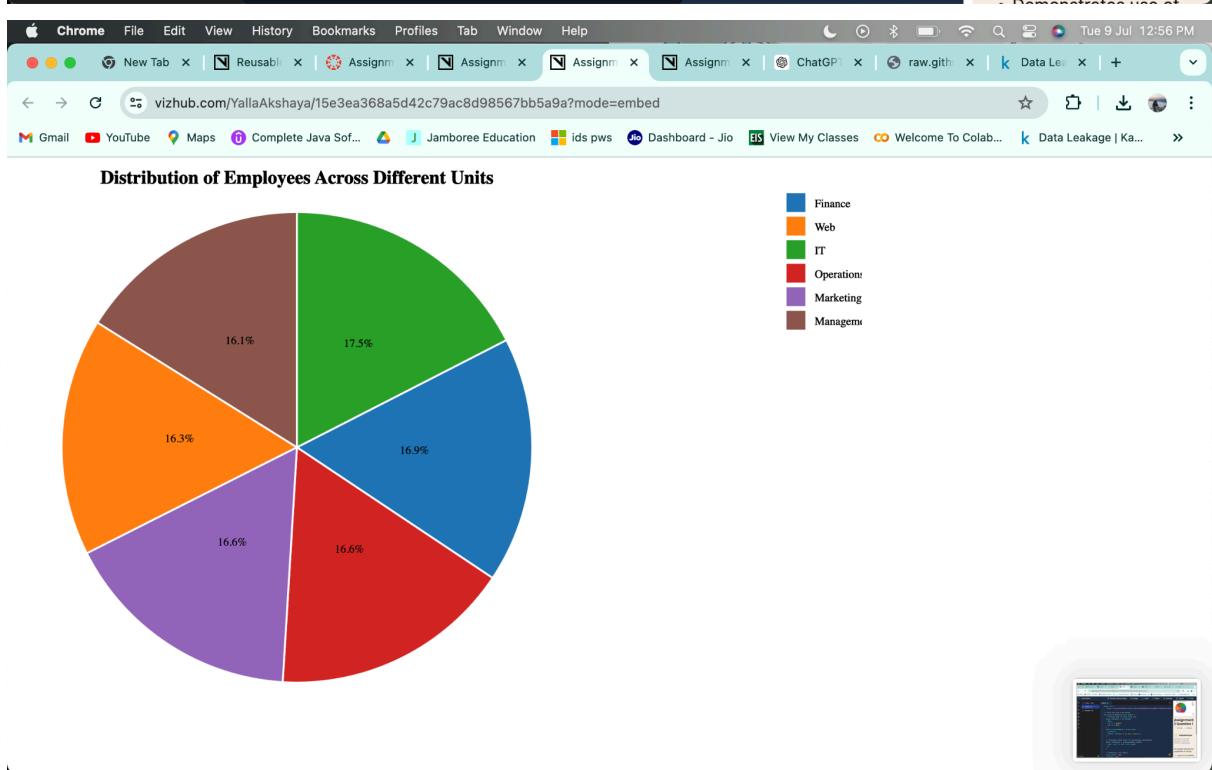
0 Forks 0 Stars

YallaAkshaya

Last edited Jul 09, 2024
Created on Jul 08, 2024
Forked from Hello VizHub

An example showing the capabilities of VizHub:

- Loads D3 via UNPKG.
- Demonstrates use of



The above are the screenshots of the code and visualization.

Understanding the Dataset:

The dataset we have used has 13 columns:

FIRST NAME: First name (Datatype: string)

LAST NAME: Last name (Datatype: string)

SEX: Gender (Datatype: character)

DOJ: Date of joining the company (Datatype: date-format)

CURRENT DATE: Current date of data (Datatype: date-format)

DESIGNATION: Job role/designation (Datatype: string)

AGE: Age (Datatype: int)

SALARY: Target variable, the salary of the data professional (Datatype: int)

UNIT: Business unit or department (Datatype: string)

LEAVES USED: Number of leaves used (Datatype: int)

LEAVES REMAINING: Number of leaves remaining (Datatype: int)

RATINGS: Ratings or performance ratings (Datatype: int)

PAST EXP: Past work experience (Datatype: int) and it has 2638 rows which gives information about different employees working in the company

Here, we represented different units with unique colours.

Understanding the Pie chart:

We have visualized the ‘Distribution of Employees Across Different Units’ for the above dataset. Different units are indicated with different colors, like

Finance -blue

Web- orange

IT- green

Operation- red

Marketing- purple

Management- brown

Also, the percentages are displayed on the chart to remove any confusion if arises.

From the pie chart, it can be observed that the employees are present in almost equal numbers in all the departments. We can also see that they are sorted in ascending order in an anti-clockwise direction. The legend is also displayed on top-right.

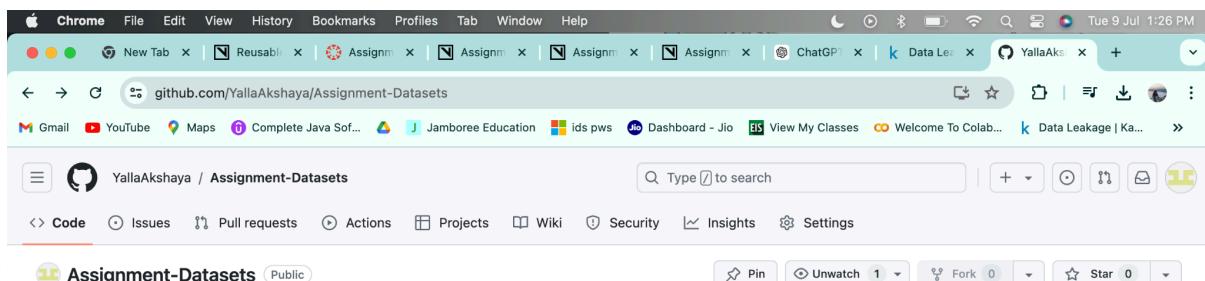
so, we can undoubtedly say that, IT unit has highest number of employees and Management unit has the lowest number of employees, However, they differ only by 1.4%

VIZHUB Link:

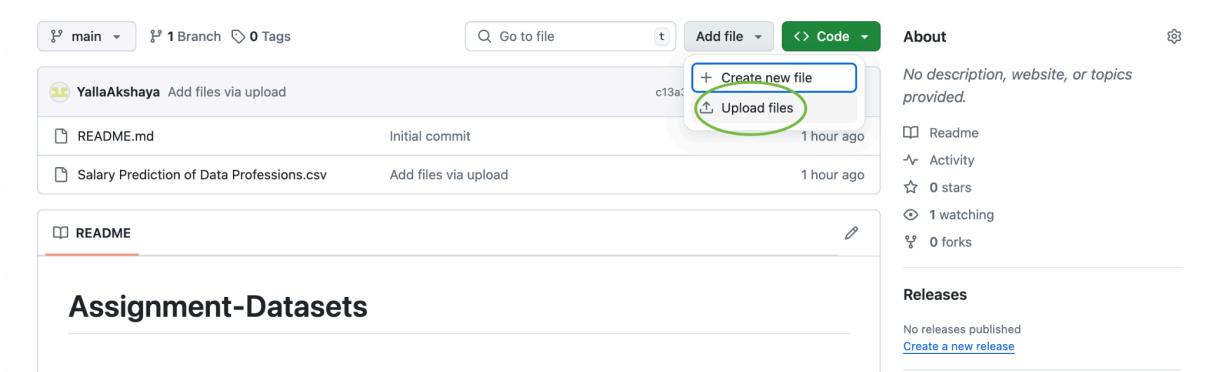
<https://vizhub.com/YallaAkshaya/15e3ea368a5d42c79ac8d98567bb5a9a?edit=files&file=index.js>

Question 2 :

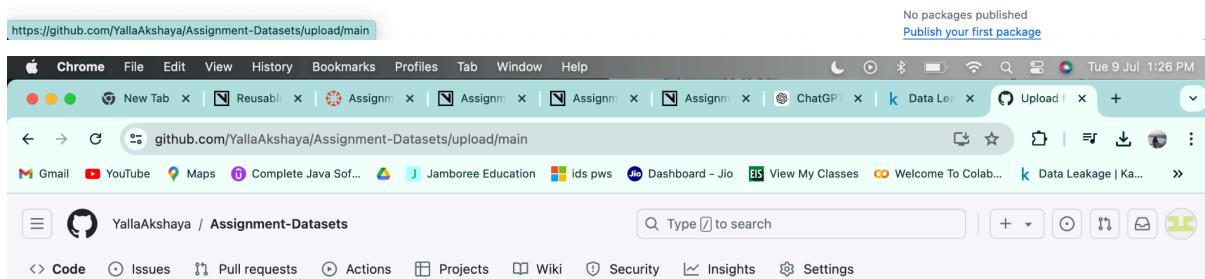
To the existing Assignment datasets repository, click on add file > upload file> choose your files > select the dataset which is AER_credit_card_data.csv which I downloaded from kaggle https://www.kaggle.com/code/dansbecker/data-leakage/input?select=AER_credit_card_data.csv



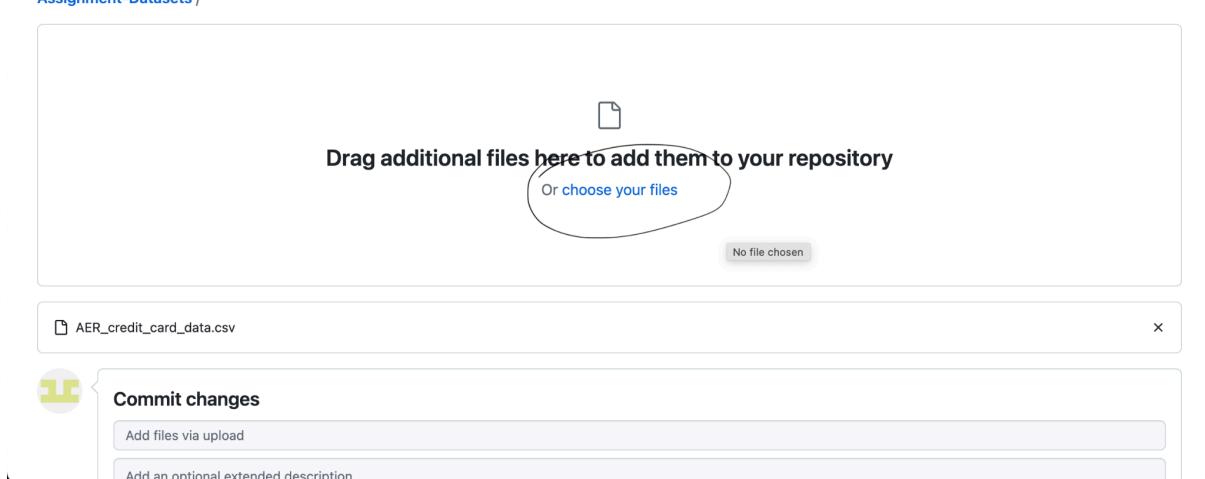
The screenshot shows a GitHub repository named 'Assignment-Datasets'. The 'Code' tab is active. A context menu is open over the file list, with the 'Upload files' option highlighted and circled in green.



The screenshot shows the same GitHub repository page, but the 'Upload' tab is now selected. The 'Upload files' button is visible in the center of the page.



The screenshot shows the GitHub upload interface. It features a large central area with a placeholder text 'Drag additional files here to add them to your repository' and a 'choose your files' link. Below this is a file list containing 'AER_credit_card_data.csv'.



The screenshot shows the GitHub commit changes interface. It displays a summary of changes: 'Add file via upload' and 'AER_credit_card_data.csv'. Below this is a 'Commit changes' section with a 'Commit message' input field containing 'Add file via upload'.

Open the dataset and click on 'raw', then it displays the whole data as comma separated values in a webpage, which can be reusable and accessible to everyone with the link.

The screenshot shows a Chrome browser window with two tabs open. The top tab is a GitHub repository page for 'Assignment-Datasets' containing the file 'AER_credit_card_data.csv'. The bottom tab shows the raw CSV data as a large block of text.

GitHub Tab Content:

Assignment-Datasets / AER_credit_card_data.csv

YallaAkshaya Add files via upload 30d5ac3 · 2 minutes ago History

Preview Code Blame 1320 lines (1320 loc) · 71.5 KB Code 55% faster with GitHub Copilot

Raw

Search this file

| 1 | card | reports | age | income | share | expenditure | owner | selfemp | dependents |
|----|------|---------|----------|--------|----------------------|-------------|-------|---------|------------|
| 2 | yes | 0 | 37.66667 | 4.52 | 0.03326991 | 124.9833 | yes | no | 3 |
| 3 | yes | 0 | 33.25 | 2.42 | 0.005216942000000005 | 9.854167 | no | no | 1,13 |
| 4 | yes | 0 | 33.66667 | 4.5 | 0.004155555999999995 | 15.0 | yes | no | 4,58,1,5 |
| 5 | yes | 0 | 30.5 | 2.54 | 0.06521378 | 30.5 | no | no | 0 |
| 6 | yes | 0 | 32.16667 | 9.7867 | 0.06705059 | 546.5033 | yes | no | 2 |
| 7 | yes | 0 | 23.25 | 2.5 | 0.0444384 | 91.99667 | no | no | 0 |
| 8 | yes | 0 | 27.91667 | 3.96 | 0.01257576 | 40.83333 | no | no | 2 |
| 9 | yes | 0 | 29.16667 | 2.37 | 0.07643376 | 137.8692 | yes | no | 0 |
| 10 | yes | 0 | 37.0 | 3.8 | 0.2456279 | 132.0 | yes | no | 0 |

github.com/YallaAkshaya/.../AER_credit_card_data.csv

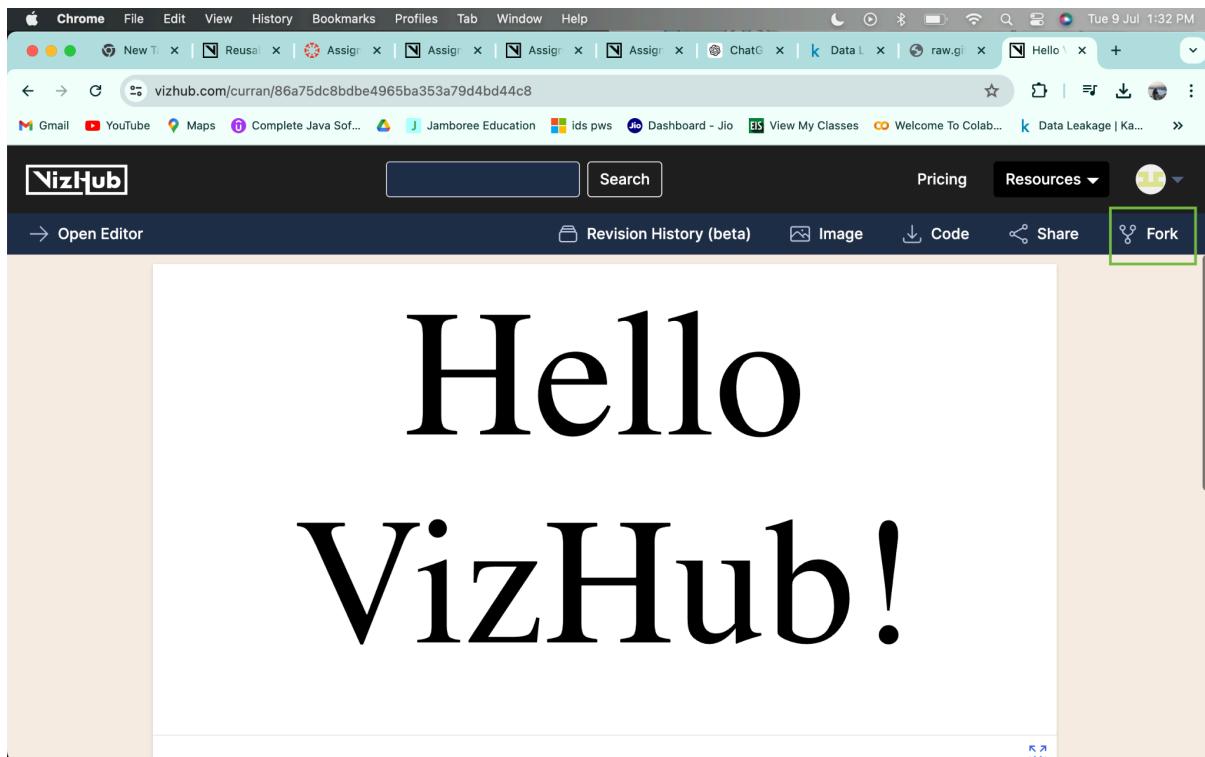
Raw Data Tab Content:

```
card,reports,age,income,share,expenditure,owner,selfemp,dependents,months,majorcards,active
yes,0,37.66667,4.52,0.03326991,124.9833,yes,no,3,54,1,12
yes,0,33.25,2,42,0.005216942000000005,9.854167,no,0,3,34,1,13
yes,0,33.66667,4.5,0.004155555999999995,15.0,yes,no,4,58,1,5
yes,0,30.5,2,54,0.06521378,137.8692,no,no,0,25,1,7
yes,0,32.16667,9.7867,0.06705059,546.5033,yes,no,2,64,1,5
yes,0,23.25,2.5,0.0444384,91.99667,no,no,0,54,1,1
yes,0,27.91667,3.96,0.01257576,40.83333,no,no,2,7,1,5
yes,0,29.16667,2.37,0.07643376,150.79,yes,no,0,77,1,3
yes,0,37.0,3.8,0.2456279,777.8217,yes,no,0,97,1,6
yes,0,28.41667,3.2,0.0197800000000002,52.58,yes,no,0,65,1,18
yes,0,30.5,3.95,0.07802456,256.6642,yes,no,1,24,1,20
no,0,42,0,1,98,0.00060606060599999999,0,0,yes,no,2,36,1,0
no,0,30,0,1,73,0.0006936416,0,0,yes,no,1,42,0,12
yes,0,28.83333,2.45,0.03879551,78.87416999999999,yes,no,0,26,1,3
yes,0,35.33333,1.9080000000000001,0.0269671,42.615,yes,no,2,120,0,5
yes,0,41.16667,3.2,0.1258194,335.435,yes,no,1,168,1,22
yes,0,40,0.03333,4,0,0.07481575,248.7192,yes,no,2,96,1,0
no,7,29.5,1,0,0.0004,0,0,2,60,1,8
yes,0,39.5,9,9999,0.06579486,548.035,yes,yes,0,28,1,0
no,3,45.75,3,4,0,0.0003529412,0,0,no,0,28,1,10
yes,0,35.25,2,35,0.02238596,43.33917,yes,no,2,115,1,1
no,1,25.16667,1,875,0.00064,0,0,no,no,2,7,0,2
yes,0,34.25,2,0,0.1311199999999998,218.52,yes,no,0,12,1,0
yes,1,35.75,4,0,0.05119225,176.6408,no,no,2,18,1,22
yes,0,42.66667,5,14,0.008949417,37.58333,yes,no,2,13,1,17
yes,0,30.25,4,506,0.1337421,502.2017,no,no,2,38,1,7
no,0,21.66667,3,84,0.0003125,0,0,no,yes,1,12,0,1
yes,0,22.25,1,5,0.0586000000000001,73.17667,no,no,0,64,1,6
no,0,34.25,2,5,0.000479999999999996,0,0,yes,no,1,12,1,0
yes,0,40,0,5,5,0.3344595999999999,1532.773,yes,no,4,74,1,19
yes,0,21.83333,2,0272,0.02532014,42.69833,no,no,0,9,0,5
yes,1,29,41667,3,2,0.15673194,417.835,no,no,0,14,1,6
no,1,24.91667,3,15,0.0003809523999999997,0,0,yes,no,1,40,1,5
yes,0,21,0,2,4663,0.2689327999999997,552.7242,yes,no,1,12,1,10
yes,0,23.83333,3,0,0.08901566,222.5392,no,no,0,12,1,5
yes,0,42.83333,3,5412,0.1834562,541.2958,yes,no,6,108,1,15
no,0,42.58333,2,2845,0.000525279,0,0,no,no,2,46,1,0
yes,0,36.58333,5,7,0.1197418999999998,568.7742,yes,no,1,36,1,16
yes,0,26.75,3,5,0.1181043000000001,344.4708,no,no,0,2,1,14
yes,0,27.75,4,6,0.1057438999999999,405.3517,yes,no,4,28,1,13
yes,0,26.25,3,0,0.124375,310.9375,yes,no,0,7,1,10
yes,0.23.33333,2,585,0.02505764,53.645,no,no,0,8,1,1
```

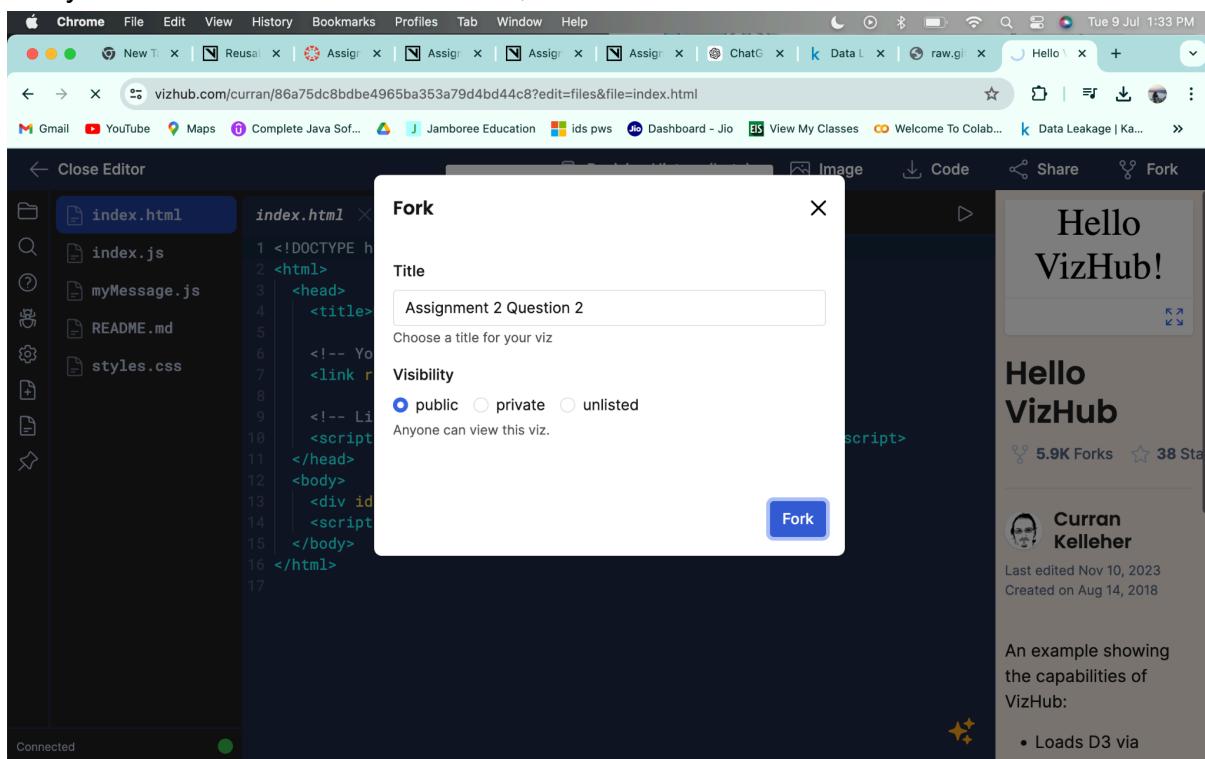
GitHub link :

https://raw.githubusercontent.com/YallaAkshaya/Assignment-Datasets/main/AER_credit_card_data.csv

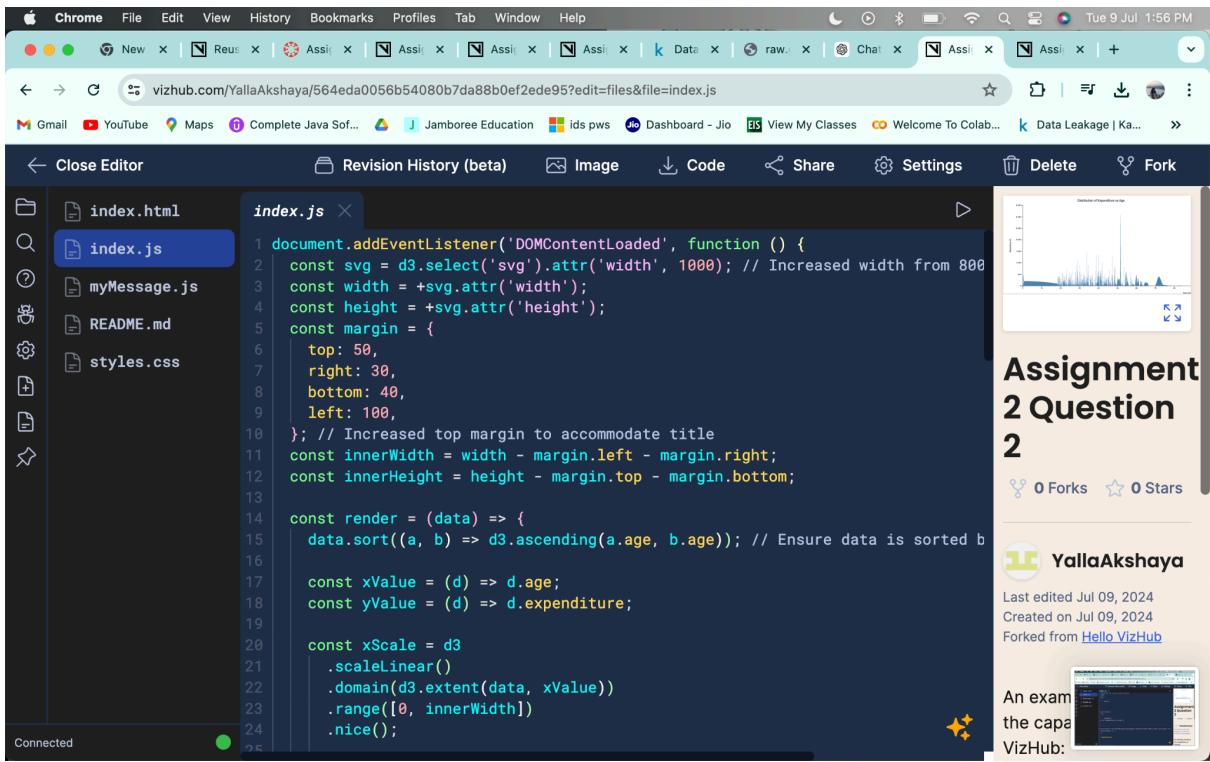
Now, click on the <https://vizhub.com/>, login to it by connecting it with your github account and select ‘hello vizhub!’ one as shown in the screenshot below.



Now, click on the fork symbol, so that you can save the project and make changes to it. Give the title and choose the visibility. Choose it to the public as it needs to be accessed by everyone who has the link. And then, click on the fork button.



Write the appropriate code to visualise the dataset using pie charts in js and html files. Making sure, that the code should work for the url of the dataset we have created using github



The screenshot shows a VizHub interface with a code editor and a preview area. The code editor displays the `index.js` file:

```

1 document.addEventListener('DOMContentLoaded', function () {
2     const svg = d3.select('svg').attr('width', 1000); // Increased width from 800
3     const width = +svg.attr('width');
4     const height = +svg.attr('height');
5     const margin = {
6         top: 50,
7         right: 30,
8         bottom: 40,
9         left: 100,
10    }; // Increased top margin to accommodate title
11    const innerWidth = width - margin.left - margin.right;
12    const innerHeight = height - margin.top - margin.bottom;
13
14    const render = (data) => {
15        data.sort((a, b) => d3.ascending(a.age, b.age)); // Ensure data is sorted by age
16
17        const xValue = (d) => d.age;
18        const yValue = (d) => d.expenditure;
19
20        const xScale = d3
21            .scaleLinear()
22            .domain(d3.extent(data, xValue))
23            .range([0, innerWidth])
24            .nice();

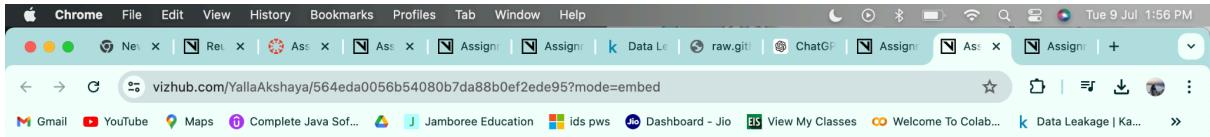
```

The preview area shows a histogram titled "Distribution of Expenditure vs Age" with the x-axis labeled "Age (years)" and the y-axis labeled "Expenditure (\$)". The distribution is highly skewed, with most values between 0 and 50 years old and a large peak around 50 years old.

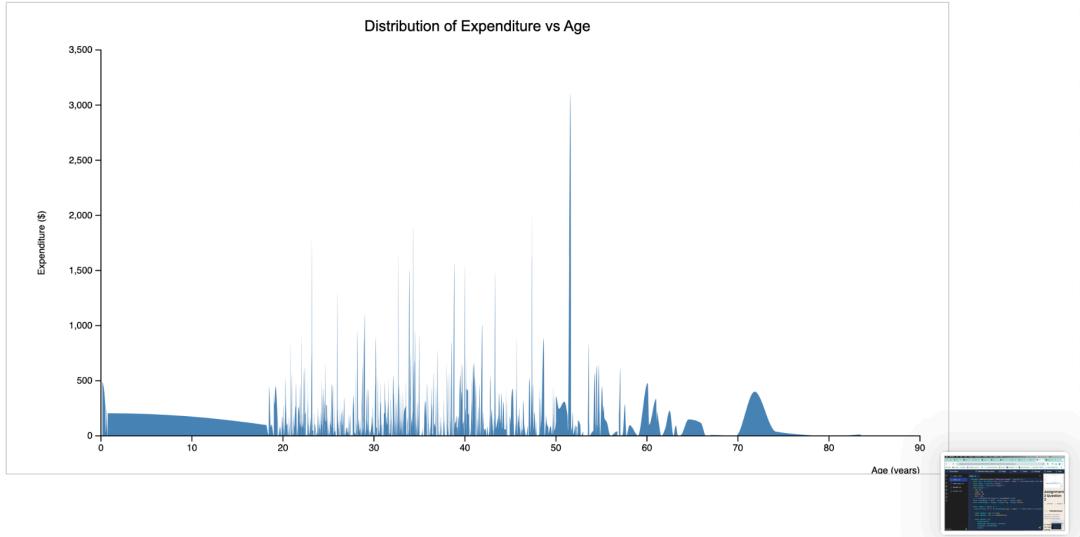
Assignment 2 Question 2

YallaAkshaya
Last edited Jul 09, 2024
Created on Jul 09, 2024
Forked from Hello VizHub

An example of the capability of VizHub:



The screenshot shows the visualization embedded in a browser tab at vizhub.com/YallaAkshaya/564eda0056b54080b7da88b0ef2ede95?mode=embed. The visualization is identical to the one shown in the VizHub interface above.



The above are the screenshots of the code and visualization.

Understanding the Dataset:

A small credit card dataset and it has 12 columns and they are:

card: Dummy variable, 1 if application for credit card accepted, 0 if not (Datatype: boolean)
reports: Number of major derogatory reports (Datatype: float)
age: Age n years plus twelfths of a year (Datatype: float)
income: Yearly income (divided by 10,000) (Datatype: float)
share: Ratio of monthly credit card expenditure to yearly income (Datatype: float)
expenditure: Average monthly credit card expenditure
owner: 1 if owns their home, 0 if rent (Datatype: boolean)
selfempl: 1 if self employed, 0 if not. (Datatype: boolean)
dependents: 1 + number of dependents (Datatype: int)
months: Months living at current address (Datatype: int)
majorcards: Number of major credit cards held (Datatype: int)
active: Number of active credit accounts (Datatype: int)

Understanding the Area chart:

We have plotted the 'Distribution of Expenditure vs Age' to visualise what is the expenditure for people with different age. This type of chart is used to represent quantitative data visually, with the area below the line filled in.

X-axis represents the **Age** (The age in years from 0 to about 90. It represents different age groups.)

Y-axis represents the **Expenditure** (The expenditure associated with each age, measured in dollars.)

Most of the expenditures are relatively low for most age groups. However, there are significant expenditure spikes at some ages.

Around 60 and 70 years old, the chart indicated higher spending, possibly due to reasons like retirement, healthcare, or travel.

The chart shows least expenditure in the very early years (close to age 0), which gradually increased.

This area chart provides insights that may be critical for industries like financial planning, healthcare, and insurance by illustrating how spending patterns change with age.

VIZHUB Link:

<https://vizhub.com/YallaAkshaya/564eda0056b54080b7da88b0ef2ede95?edit=files&file=index.js>

Please refer to the above vizhub link for any doubts.