Stats 306, Winter 2024, Final Group Project

Your group number and each team member names

Throughout this course, you've dedicated yourself to refining your analytical abilities using R programming language. These skills are highly coveted in today's job market! Now, for the semester project, you'll apply your learning to craft a compelling `Data Story` that can enrich your portfolio and impress prospective employers. Collaborating with a team (up to 5 members of your choosing), you'll construct a Data Story akin to the example provided here: https://ourworldindata.org/electric-car-sales#article-licence

Data is already in the `data` folder. This data is downloaded from: https://www.iea.org/data-and-statistics/data-tools/global-ev-data-explorer

You'll conduct Exploratory Data Analysis (EDA) on the provided datasets. Although the provided article already includes 8 diagrams for you to replicate initially, I'm also interested in seeing how each team will expand upon the initial analysis and generate additional insights and charts.

Deliverable

1. Requirement-1 (1 pt)

You should show at least 2 steps you adopt to clean and/or transform the dataset. The opportunities include dropping some columns that are irrelevant. For e.g., in the ev-historical.csv data, does the column 'historical' have any significance? Find at least 2 such steps you would take to clean up your data to proceed further with analysis.

2. Requirement-2 (1 pt)

Are there any opportunities to make to use `pivot_wider` or `pivot_longer` to make the data 'tidy'? Explain why or why not this data is a candidate for applying tidy transformations.

3. Requirement-3 (10 pt)

For the EDA part, in addition to the 8 diagrams (4 pt) already given in the article, you need to plot 6 (6 pt) more diagrams to show correlations, frequencies, and/or relationships between various variables with plots of 5 different types (bar, line, heatmap, facet, etc.). Every plot should have a title and the x/y axis should have legible labels without any label overlaps for full credit. Provide a summary of your interpretations from the plots after each one.

4. Requirement-4 (2 pt) By this phase, you have a pretty good understanding of your data. Now, you will apply predictive analytics by building a suitable ML model to make some predictions by selecting suitable independent/dependent variables from the dataset. Explain how you narrowed down on the variables chosen for building your model and make some future predictions on some aspect of EV.

5. Requirement-5 (1 pt)

You should have a conclusion, highlighting the main insights that you were able to derive from your analysis. I will be particularly interested to know what else you are able to find beyond what the author of the article discovered.

6. Extra Credit (1 pt) You may build an interactive shiny app to show your analysis instead of submitting your project as a PDF and .Rmd file.

Submission

- You will upload the zip file containing finals.Rmd file and its PDF as a deliverable to Canvas.
- You will present your findings by creating a video; all team members should explain their part in the project to receive credit. You will share the URL of the video on Canvas for us to evaluate.

It is not necessary to prepare slides (if you do it doesn't hurt) for the presentation. You may speak by showing the diagrams and/or code from your Posit project. Every team member should explain their part in the project along with the insights they derived by explaining the charts and summaries for full credit to each member.

Your project will be evaluated for its meaningful/insightful EDA and predictions.

Note:

- Each plot must be accompanied by a summary that clarifies the rationale behind its creation and what insights the plot unveils. Every diagram should possess standalone significance, revealing something more compelling than the other charts
- After the deadline, instructors will select the top three outstanding analytics projects. The teams responsible for these exceptional analyses will have their video shared with the class

```
library(tidyverse)
df1 <- read_csv('data/ev-historical.csv')
```

```
## Rows: 2776 Columns: 8
## -- Column specification --------------------------------------------------------
## Delimiter: ","
## chr (6): region, category, parameter, mode, powertrain, unit
## dbl (2): year, value
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```
df2 <- read_csv('data/ev-steps-projected.csv')
```

```
## Rows: 270 Columns: 8
## -- Column specification --------------------------------------------------------
## Delimiter: ","
## chr (6): region, category, parameter, mode, powertrain, unit
## dbl (2): year, value
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```