Telling Stories with Data Dashboard

Jacob Clement

Inital submission

The final product for my first submission is `scripts\price_mileage_6.py`.

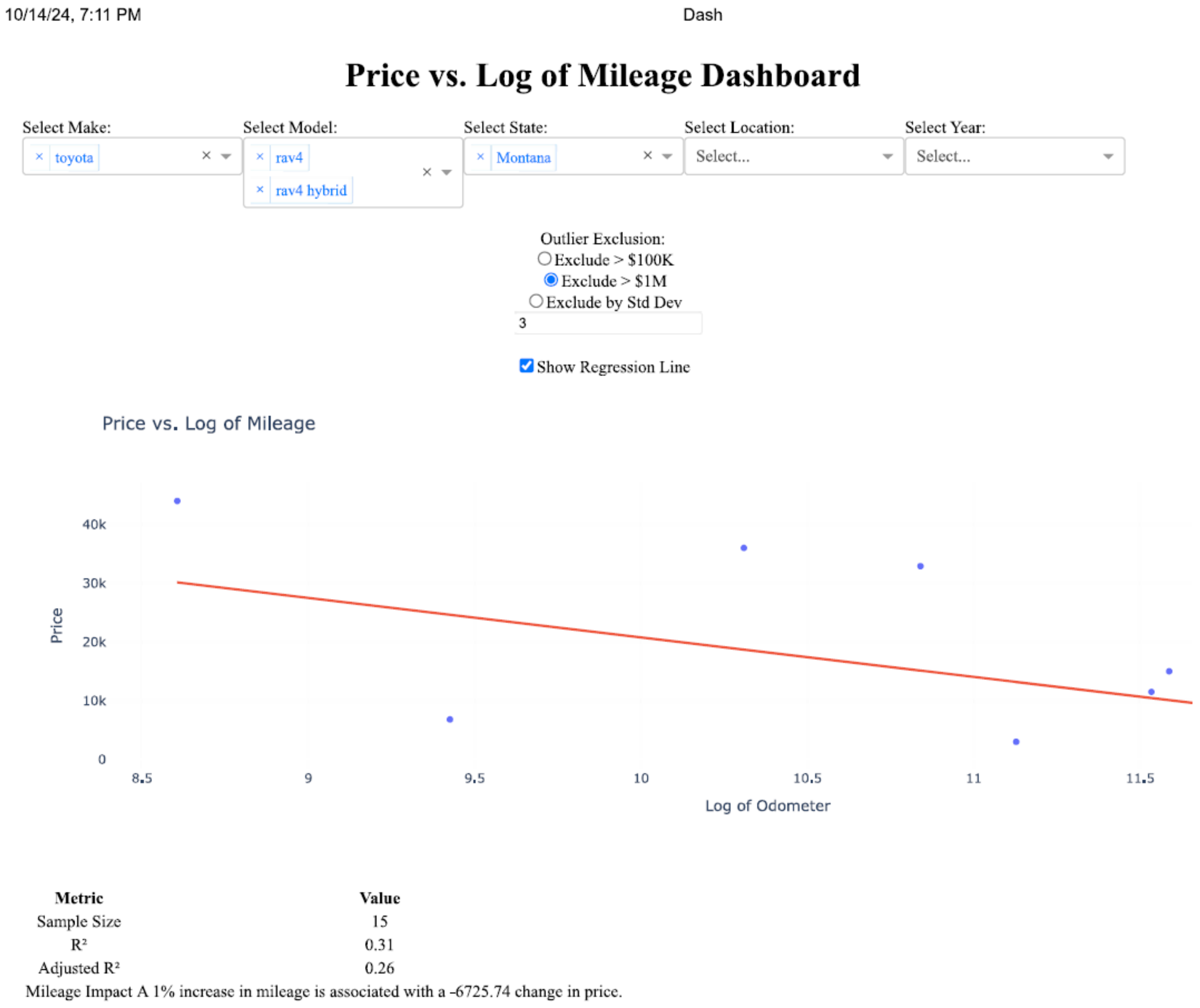
I scraped `state` from craigslist and was almost done with the process so I figured I would submit what I have and try to work with the data for Montana to create a final polished product.

ChatGPT: dashboard creation: https://chatgpt.com/share/670db7df-a534-800a-b24e-4a4fed4b777c

dashboard debugging: https://chatgpt.com/share/670db7df-a534-800a-b24e-4a4fed4b777c

Web Scraping: https://chatgpt.com/share/670daba1-86f0-800a-9c97-b771790225d8

Here is a static view of my car, I stopped filtering as the sample is small. My goal is to include a version of this dynamic scatterplot in my final dashboard.



10/14/24, 7:11 PM Dash

Price vs. Log of Mileage Dashboard

Select Make: toyota
Select Model: rav4, rav4 hybrid
Select State: Montana
Select Location: Select...
Select Year: Select...

Outlier Exclusion:
○ Exclude > $100K
● Exclude > $1M
○ Exclude by Std Dev
3

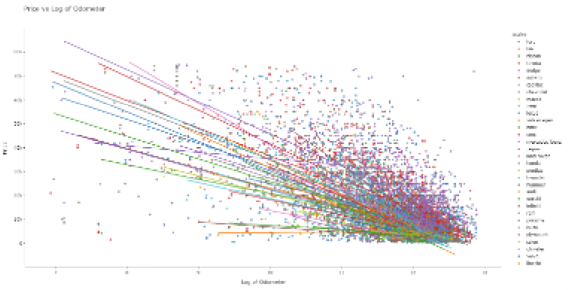
☑ Show Regression Line

Metric	Value
Sample Size	15
R²	0.31
Adjusted R²	0.26
Mileage Impact	A 1% increase in mileage is associated with a -6725.74 change in price.

Github Repo: https://github.com/jacobchrono/telling_stories_dashboard

Second Submission

You should use `html\dynamic_scatter_fail.html` for future examples of bad graphs. ChatGPT defended its output but I could not attach the full convertsation due to a pasted image.



what do you think of this graph chat gpt?

This scatter plot looks quite comprehensive and visually rich. Here's a quick review of its strengths and potential improvements:

What Works Well:

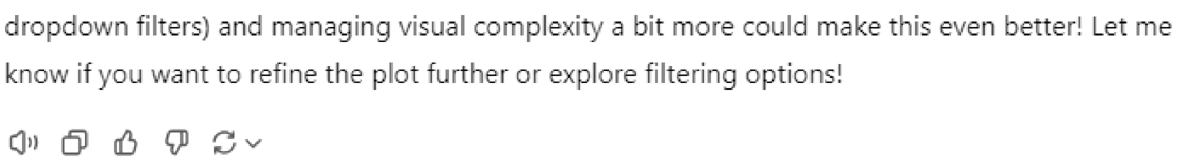
1. **Color-coded by Make:**
   Different makes are distinctly colored, which helps in visually clustering points by brand.
2. **Regression Lines by Make:**
   Multiple regression lines offer insight into the trend of prices versus odometer readings for each car make.
3. **Log Transformation of Odometer:**
   Using the log of the odometer effectively handles skewed data, making the plot more interpretable.
4. **Minimalistic Theme:**
   The graph is clean and not cluttered with unnecessary gridlines or annotations.

Areas for Improvement:

1. **Overlapping Data Points:**
   Some points overlap, making it difficult to identify distinct clusters. Consider **transparency adjustments** or **jittering** to reduce overlap.
2. **Too Many Colors:**
   While color is useful, with so many makes, it may be visually overwhelming. Grouping **similar car brands** or allowing interactive filtering could improve clarity.
3. **Interactive Dropdown or Button Options:**
   Adding interactive filters for **Make, Model, Transmission, or Drive** would allow users to focus on specific data and reduce clutter.
4. **Axis Label Formatting:**
   Add **currency formatting** to the y-axis to emphasize the price units.
5. **Legend Management:**
   With a long legend on the right, it could be helpful to make the legend scrollable or collapsible in an interactive plot to avoid taking up space.

Summary:

This is a great start with solid insights from the regression lines. Implementing interactivity (such as



dropdown filters) and managing visual complexity a bit more could make this even better! Let me know if you want to refine the plot further or explore filtering options!

My vision is to add a widget here the user can build a custom regression model of the full data. I had to limit the number of filters for file size and ability to share so my idea my not be within the scope of the project. Can we share a hosted version over the message board? The final product for this submission is `html\vehicle_dashboard.html`. The relevant python script is `scripts\combine_montana_4.py.`

https://chatgpt.com/share/670ea0a5-5724-800a-8e8f-397db08135d9