Project idea brainstorming

* Social media marketing (for singers/musicians?)
  + Group by region, age, genre
  + Recommend hashtags/keywords, could present as a word cloud
  + Use LangChain to augment the results
  + Would need a dataset
  + Also need to figure out what kind of DS problem this is (classification or clustering make some sense, but if the user is picking the specific group they belong to, it’s not clear how this would apply?)
* I’m not sure any of the stuff I (Geoff) usually work on is going to work for this at the moment, but if anyone wants to check for some ideas:
  + Most of the recent work I’ve been doing has drawn from <https://openalex.org>
  + This usually involves downloading data using the APIs and formatting it into our own datasets (often in combination with data from elsewhere).
  + Pre-print of one of the papers I worked on (accepted for publication): <http://arxiv.org/abs/2306.16554>
* ~~In addition to the Government of Canada’s open data site for government data, they also run a new Open Science Data Platform (~~[~~https://osdp-psdo.canada.ca/dp/en~~](https://osdp-psdo.canada.ca/dp/en)~~), and there might be some interesting datasets there.~~ OSDP seems to be offline right now.
* There’s also an open data page for Halifax at <https://data-hrm.hub.arcgis.com/pages/open-data-catalogue> (though a lot of this is GIS – mapping – data).
  + There’s lots of traffic & transit data here.

# (Ye Wang) Social media trends data analysis

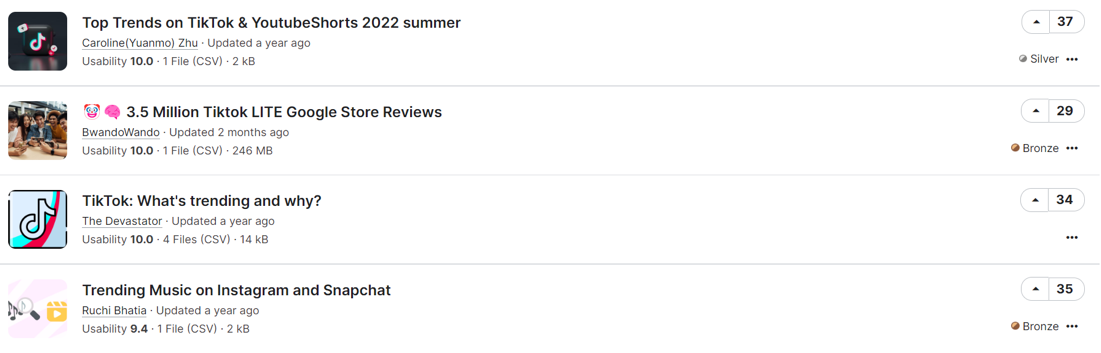
**Reason: Easier to meet professor’s requests. Clean Data is easy to get. All feature of social media and products are reasonable and easy to understand, which means no hidden relations of features need to be found.**

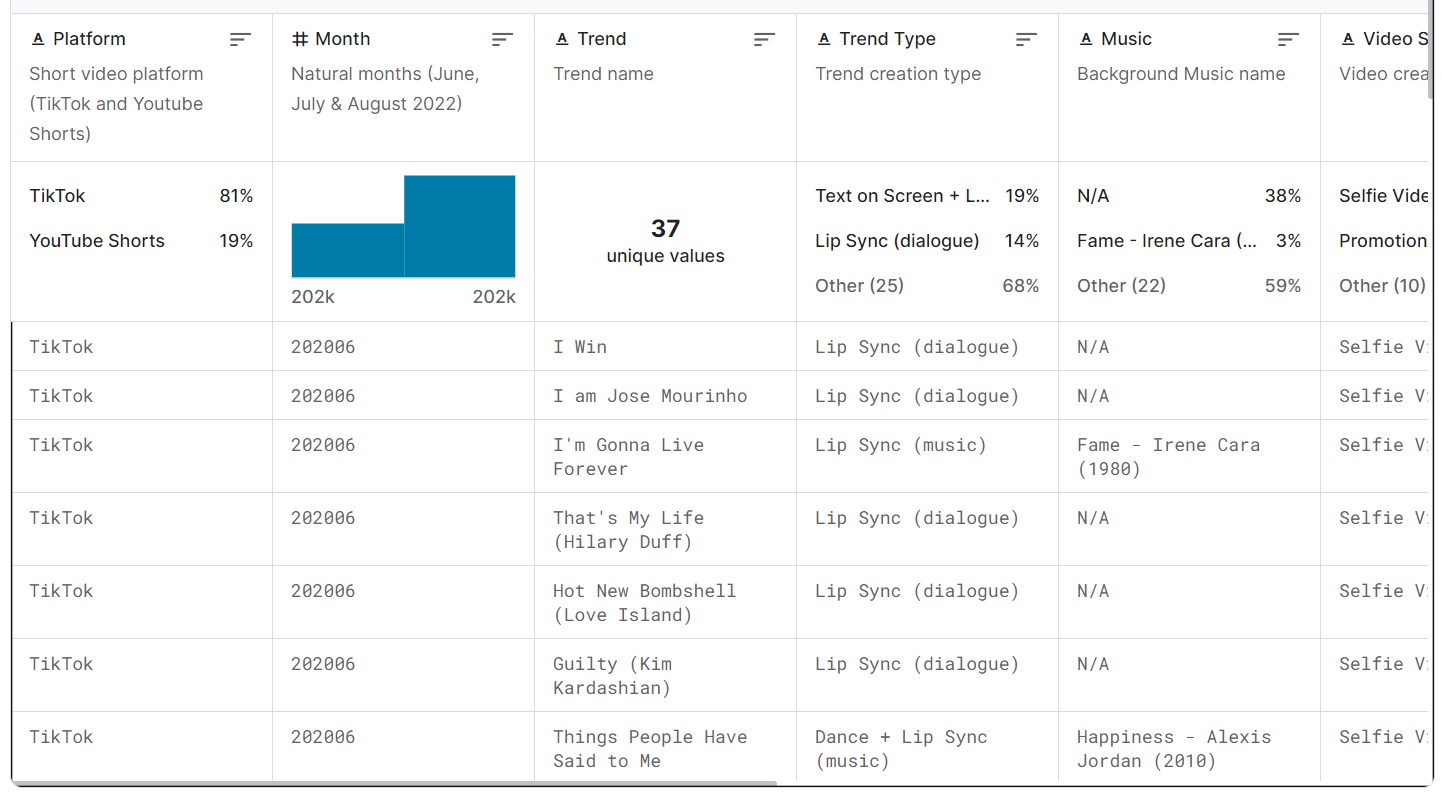
**Main goal:** Analyze how the trend changes through time and see how these changes affect the actual marketing strategy for products.

**Theories included: Classifications (about features), Regressions (about changes)**

**Data source:** Can be widely found from Kaggle, or social media’s platforms. Very easy to get.

<https://www.kaggle.com/datasets>





And they are all cleaned and organized, waiting for us to use. Very convenient. Easy to complete.

**Possible features/variables:**

**Categorical** – Hashtags, areas (country, culture...), content/trend type (music, dance, platform (TikTok, YouTube, Instagram...), meme, and others

**Quantitative** – Users age, likes, shares, comments, video length, and others

**Helpful tools:** Jupiter, langChain, R, and others

# Electrical Cars:

Source：

<https://catalog.data.gov/dataset/electric-vehicle-population-data>

Reasons: 1, Open raw government data set. Fit the request of professor. 2, Future trend. 3, Features are clear and easy to understand. 3, Sudhan love cars and knows a lot of things about it.

Details:

<https://data.wa.gov/Transportation/Electric-Vehicle-Population-Data/f6w7-q2d2/about_data>

Possible goals:

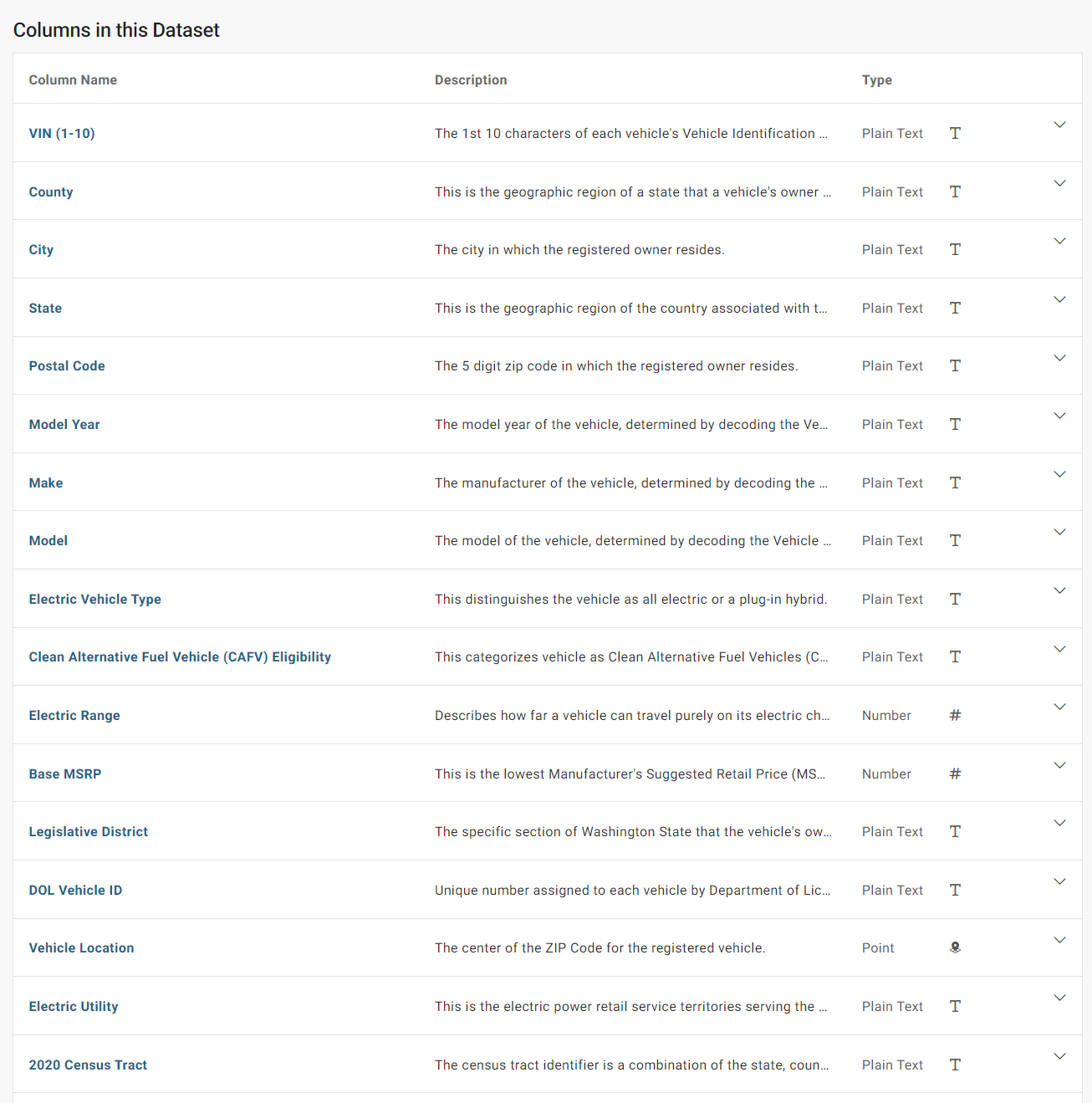
1, Market Analysis: Understanding the distribution and popularity of electric vehicles by make, model, and year in different regions. This can provide profit insights for multiple business purposes.

2, Trend Analysis Over Time

Analyze how the adoption of electric vehicles has changed over time by examining the distribution of model years. This can provide insights into the growth rate of electric vehicle usage and how consumer preferences have evolved.

3, Geographic Analysis: Understanding the geographic distribution of electric vehicles, which can be useful for city planning, infrastructure development, and targeted marketing strategies. Providing future deployment insights both for companies and government apartments.

Features:



# References we can use later:

Government of Canada.(2023,December 19) Canada’s Electric Vehicle Availability Standard (regulated targets for zero-emission vehicles)

<https://www.canada.ca/en/environment-climate-change/news/2023/12/canadas-electric-vehicle-availability-standard-regulated-targets-for-zero-emission-vehicles.html>

Afandizadeh, S., Sharifi, D., Kalantari, N. et al. Using machine learning methods to predict electric vehicles penetration in the automotive market. Sci Rep 13, 8345 (2023). <https://doi.org/10.1038/s41598-023-35366-3>

Wang, F.-K., Chang, K.-K. & Tzeng, C.-W. Using adaptive network-based fuzzy inference system to forecast automobile sales. Expert Syst. Appl. 38, 10587–10593 (2011).

Hülsmann, M., Borscheid, D., Friedrich, C. M. & Reith, D. General sales forecast models for automobile markets and their analysis. Trans. Mach. Learn. Data Min. 5, 65–86 (2012)

Kitapcı, O., Özekicioğlu, H., Kaynar, O. & Taştan, S. The effect of economic policies applied in Turkey to the sale of automobiles: Multiple regression and neural network analysis. Procedia Soc. Behav. Sci. 148, 653–661 (2014)

Bas, J., Zou, Z. & Cirillo, C. An interpretable machine learning approach to understanding the impacts of attitudinal and ridesourcing factors on electric vehicle adoption. Transp. Lett. 15, 30–41 (2023).

Zhang, Y., Zhong, M., Geng, N. & Jiang, Y. Forecasting electric vehicles sales with univariate and multivariate time series models: The case of China. PLoS ONE 12, e0176729 (2017).

Saxena, P., Bahad, P. & Kamal, R. Long short-term memory-RNN based model for multivariate car sales forecasting. Int. J. Adv. Sci. Technol. 29, 4645–4656 (2020).

Beggs, S., Cardell, S. & Hausman, J. Assessing the potential demand for electric cars. J. Econom. 17, 1–19 (1981).

Calfee, J. E. Estimating the demand for electric automobiles using fully disaggregated probabilistic choice analysis. Transp. Res. Part B Methodol. 19, 287–301 (1985).

**Useful related data sets:**

<https://www12.statcan.gc.ca/census-recensement/2021/dp-pd/prof/details/page.cfm?Lang=E&SearchText=Ontario&DGUIDlist=2021A000235&GENDERlist=1,2,3&STATISTIClist=1&HEADERlist=0>