Stellantis 2025 Y-Data Project Scoping

\*\* As I see it (and would like to further discuss it with you) we can:  
1. Engineer driving behavior features and add them to current models.

2. Use those features to first cluster drivers and then add the driving style label as a feature to the models.

The following document is written in light of option 2. I guess that the data can help us choose between the options.

Intro:

* Project Description:
  + Enhance Stellantis' ETA and energy consumption prediction models by integrating driving behavior patterns derived from vehicle sensor data.
* Milestones:
  + Initial Presentation (31/3): Problem statement, dataset, work plan, and initial insights.
  + Final Presentation and model delivery (31/7).
* Team:
  + **Students**: Alexandra Braginsky, Shay Ofir-Geva, and Yonatan Goldstein.
  + **Mentor**: Oren Elisha.

Product:

* **Stellantis's main contact**: Amir Baruch.
* **Client’s pain**:
  + Current ETA and energy consumption predictions do not incorporate driving behavior, potentially reducing accuracy.
  + Inaccurate predictions affect user satisfaction and route planning efficiency.
* **Model Outputs**:
  + Driving‐style classifier capable of differentiating driving behavior patterns.
  + ETA and end‐of‐trip energy state (e.g., battery SOC) predictions, leveraging the driving‐style output as an additional feature.
* **Deployment**:
  + Stellantis' vehicle navigation or telematics systems (out of scope for this project).
* Business KPIs:
  + Significant clustering of users according to their driving behaviors.
  + Significantly Increased accuracy of ETA and energy consumption predictions, relative to current baseline predictions, when adding the driving‐style output as an additional feature to the models.
  + Improved user satisfaction with ETA and energy consumption predictions (will not be evaluated in this project).
* **Business Impact**:
  + Higher user satisfaction with vehicle navigation and telematics systems.
  + Enhanced trust in Stellantis-brand vehicles.
* Data:
  1. Time period
  2. Amount of data
  3. Population
  4. Labels
  5. Train / test sets

Solution:

* **Baseline**:
  + Existing ETA and energy consumption models without driving pattern features.
* **Initial modeling approach**:
  + **Feature Engineering**: Create features encompassing driving behaviors.
  + **Clustering**: Cluster users based on their driving behaviors.
  + **Integration**: Incorporate driving style labels into existing ETA and consumption models.
* **Solution assumptions and risks:**
  + 1. What are the product / data assumptions?
    2. What is the risk from these assumptions?
    3. Risks mapping (research / model / data, engineering, business)
* **Evaluation metrics**:
  + Silhouette Coefficient and domain‐oriented checks of relevance for driving style clustering.
  + Mean Absolute Error (MAE) and Root Mean Squared Error (RMSE) for ETA and Energy‐Consumption Predictions.

1. Requirements section
   1. Dataops
      1. labeling & data curation efforts
      2. Datadev needed to get / process the data
   2. Engineering requirements
      1. Backend / frontend dev need to integrate or deploy the model
      2. Model consumption and API
   3. A/B test - notes / requirements
2. Legal consult requirements
   1. Data usage that need legal approval (in terms of GDPR, controller-processor relation, PII, 3rd party data, open datasets)
   2. Open source / packages in use (verify licencing etc)
   3. AI regulations