Individual Work Log

PROJECT NAME:	OBD-II Based Predictive Maintenance System		
STUDENT NAME:	**Dang Khoa Le**		
STUDENT ID:	**103844421**	**WEEK # (& dates covered):**	**#7**

TASKS	STATUS	TIME SPENT	ACTION ITEM/NOTE
Task	**Status**	**Time Spent**	**Action Item/Note**
Collect and explored raw OBD-II datasets (logged in CSV) with Python pipelines	Completed	2.5 hours	Reviewed structure, missing values, and key sensor variables from merged raw CSVs (KIT, Toyota Etilos, LEVIN)
Research on raw OBD-II data (Hex-based) and solutions to convert them into interpretable format	Completed	2 hours	Raw OBD-II data from ECU is hex-based, not a human readable and ML-friendly. Use either OBD Raw Data Parser or CAN Decoder to decode and transform to proper formatting.
Performed deep data cleaning operations	Completed	2 hours	Fixed encoding issues (e.g., Â, °C),handling non-string features, forward-filled missing + corrupted values, removed extreme outliers
Feature Engineering + Merging CSVs	Completed	2 hour	Merged grouped and identical schema CSVs (simulating different log attempts) with origin tagging; engineered derived features like trip delta, throttle shifts
Conducted EDA + Insight Summary Reporting	Completed	2 hours	Statistical profiling, anomaly flags, and pattern discovery (e.g., RPM + Temp correlation); prepared markdown reports
Created Colab Notebook walkthrough	Completed	1.5 hours	Structured a full notebook guiding data loading, cleaning, EDA, and interpretation steps
Prepared Powerpoint slides and summary for client presentation	Completed	1 hours	Explained workflow: cleaning, merging, and insight extraction for non-ML-ready datasets
Documented real-world data collection, processing and training challenges	Completed	1 hour	Highlighted limitations of sensor saturation, timestamp gaps, and vehicle-specific inconsistencies
TOTAL WEEKLY TIME SPENT		**14 hours**	

TASKS PLANNED FOR NEXT WEEK	EXPECTED COMPLETION
Get OBD-II on Monday and start data collection	Week 8
Label data using heuristics from DTC and engine indicators	Week 8-9
Define new constraint with data collection, interpretation and cleaning procedures	Week 8-9
Having discussion on data science tasks with the unit and client.	Week 8

Summary/weekly reflection for Week 7:

- **Key Tasks Done:**

Explored and profiled merged OBD-II datasets as well as researching on raw hex-based data logs and solutions, with a focus on structure, nulls, and data readiness. Performed cleaning routines: fixed corrupted encodings, filtered outliers, and handled low-level missing data. Merged and tagged grouped CSVs for consistent modelling structure. Led exploratory feature engineering to capture trip-level and engine-load patterns. Created a markdown and presentation summary for team/client.

- **Key Learning:**

Real-world telemetry data requires robust cleaning pipelines to be usable for ML. Many automotive sensors max out at fixed values or are missing in older vehicles. Data types, naming conventions, and units vary across sources and require harmonization. Exploratory statistics surface useful ML features (e.g., RPM, throttle, engine temp).

- **Literature/Resources Reviewed:**

- ELM327 OBD-II standard documentation and vehicle PID coverage references.
- Raw CSV-based KIT, Toyota Etios, LEVIN OBD-II datasets. Techniques for merging heterogeneous sensor files and aligning time series across vehicle trips.
- Raw hex-based dataset (ECUPrint Dataset, VED – Vehicle Energy Dataset). Techniques to decode into CSVs for interpretation.

- **Issues Faced:**

Sensor readings like MAF and battery voltage were capped (likely placeholder values). Missing timestamps in some rows limited temporal alignment. Unit inconsistencies (e.g., °C vs. C, string-based accelerometer data) added to preprocessing effort. Some features required feature engineering from multiple fields.