Progress Report 1

Name: Dundee Adriatico, 300393449

Project: Riipen EMS Data Analysis and Forecasting

Repo: https://github.com/dondemici/W26_4495_S2_DundeeA/

Work Log

Week	Date	Number of Hours	Description of Work Done
1	18-Jan-26	0.5	Meeting with Priya, initial discussion of topics
	21-Jan-26	2	NEMSIS Data initial proposal
	23-Jan-26	0.5	Meeting with Solaris, Tony Tsui
	24-Jan-26	1	Initial works on the final proposal, intro and proposed research project
	25-Jan-26	2	Align Intro and Proposed Research Project on the final proposal – Riipen link finally given
	26-Jan-26	4	Final works on the final proposal – Estimate and update timelines, build contract, update references, update AI usage table and prompts and perform final review and revisions
2	30-Jan-26	2	Review of NEMSIS website, submission of data request for the project, communication of updates to Tony Take Datacamp course on Introduction to Databricks, Setup a community edition Databricks account,
	31-Jan-26	2.5	Review data sent by NEMSIS, extract data (about 2 hrs. processing to extract 200GB data), build initial python code to review data
	03-Feb-26	1.5	Update project plan to Agile format, reconsider performing a subset of data performing activities of the whole project in a shorter span of time, create ems eda notebook
	05-Feb-26	1.5	Provide updates to Riipen stakeholder, further revise project plan. Update ems eda notebook
	06-Feb-26	1.5	Provide updates to Priya, consolidate and review all generated files including notebook, upload to github
	08-Feb-26	2.5	Build end-to-end notebook pipeline, troubleshoot time data format, review next steps for the succeeding week.

Summary Description

This week I focused on turning the NEMSIS data into a usable EMS demand prototype while refining the project plan and tool setup. I moved from initial exploration into a working modeling pipeline and clarified how Jupyter and Databricks will be used going forward.

- Reviewed NEMSIS documentation and submitted the official data request, clarifying how key variables can support EMS demand and injury-risk modeling.
- Set up Databricks Community Edition, completed an introductory Databricks course, and began exploring the large NEMSIS extract with an EMS EDA notebook.
- Converted the initial project plan into an Agile/sprint format, defined a smaller "vertical slice" focused on forecasting high-strain EMS demand from a subset of the data, and provided progress updates to Riipen stakeholders and the course supervisor.
- Built an end-to-end Jupyter notebook pipeline that cleans and renames key NEMSIS fields, parses available time data, engineers time/scene/patient mix features, aggregates incidents to hourly call volume, and trains baseline regression and high-strain classification models.
- Attempted to port this pipeline to Databricks, encountered timestamp and type-casting issues (e.g., "Not Applicable" and "." values), and iteratively revised code with synthetic time indices and more robust casting to understand platform constraints.
- Adjusted the short-term plan so that Jupyter remains the main environment for model development on the subset data, while Databricks is used primarily for ETL and building clean hourly aggregated tables that can later support scalable, scenario-based EMS demand forecasting.

Overall, the project has moved from planning and setup into concrete modeling work, surfaced important data-quality and platform issues early, and produced a functioning prototype pipeline that will guide the next round of feature refinement and larger-scale deployment.

Repo Check-in

I have committed my work regularly to the shared GitHub repository, focusing on the initial EDA and the first end-to-end modeling prototype. The main implementation artifacts are under the Implementation directory, and documentation is stored under Documents and Reports. ie proposal and this progress report.

Folder File	File	Brief Description
DocumentsAndReports	DundeeA_Proposal.pdf	Original project proposal outlining goals (EMS high-strain demand forecasting and injury-risk context), data sources (NEMSIS), methods, and initial timeline.
	DundeeA_ProgressReport1.pdf	Current progress report summarizing work completed since the proposal, issues encountered (data quality and Databricks constraints), adjustments to the plan, and the updated short-term timeline.
Implementation	ems_eda.ipynb ems_eda2.ipynb	Initial exploratory notebook for the NEMSIS subset: inspects schema, value distributions, missing/coded values (e.g., "Not Recorded"), and highlights candidate variables for EMS demand and injury-risk modeling.
	ems data review.ipynb ems_data_review databricks.ipynb	End to end Jupyter pipeline for a vertical slice: renames key NEMSIS columns, parses available time fields, engineers time/scene/patient mix features, aggregates incidents to hourly workload, and trains baseline regression and high strain classification models.
	nemsis_sprint0_3_sample.csv	Initial consolidated scoped sub dataset for exploration, sourced from the full 200GB NEMSIS dataset.
	NEMSIS Header Table.xlsb	Description of NEMSIS header to make EDA meeaningful
	sample_DISPATCHDELAY.txt sample_Pub_PCREvents.txt sample_TURNAROUNDDELAY.txt	Initial raw dataset sample selected from NEMSIS full dataset

These check-ins ensure that both the exploratory code and the evolving documentation are tracked and available for review as the project moves into more advanced modeling and forecasting.

AI Use Section

AI Tool Name	**Version Account Type**	**Specific Feature for which the AI Tool was Used**	**Value Addition** *(What value did you add over and above what AI did for you?)*
Perplexity	Education Pro	Asked AI to surface literature that would support framing the problem that demonstrates occupational hazard encountered EMS personnel	Having done a coop role in VCH, I learned that work-related injuries are handled by WorkSafeBC so I also asked AI to check for any statistics coming from the organization, this is on top of the 2 it produced on CUPE and JEMS
Perplexity	Education Pro	Asked AI to confirm if there's any AI or ML related studies or projects published using the NEMSIS data in order to check uniqueness of my study	Gave input to AI about the focus of my studies which is really about forecasting work-related injuries for emergency personnels so that it serves as comparison for other studies that may have been done
Perplexity	Education Pro	Asked AI to evaluate my timeline if reasonable and for any crucial activities I may have missed	Provided the milestones I crafted as well as the activities envisioned so that AI provide further input on anything I missed
Perplexity	Education Pro	Asked AI to provide crash course training for databricks	Provided level of knowledge as well as planned usage so that I can get a focused kind of training
Perplexity	Education Pro	Asked AI to troubleshoot the errors from the codes created to be efficent	Provided context of the code as well as the exact error encountered and section or line it was encountered.

List of AI Prompts

Prompts
Given my project and objective on analyzing EMS data and forecasting information to improve QA protocols to lessen work-related injuries of emergency personnel, find relevant literature that would support demonstrating the occupational hazards encountered by these workers.
When the literature was given, I further prompted if AI can give me stats by WorkSafeBC, having learned about the organization during my coop role time in VCH.
Given that I am doing a project on NEMSIS data, provide confirmation on any studies done in the past that is related to AI or ML so that I understand the significance and uniqueness of my study. My study focuses on forecasting work-related injuries for emergency personnels.
Review and evaluate my timeline of the project from EDA, cleaning, model evaluation and selection until dashboard reporting and check if reasonable given the 10-12.5 hrs I can allot weekly for a 2-month period. Mention any activities I may have missed to include.
Provide a step-by-step procedure on how to import data in databricks. Include steps on how to setup the needed data or database including steps on how to create a notebook.
Given the following timeframe code and the equivalent error encountered, what areas do I need to look out for in order to solve the error.