"Telematics Unleashed: The Truth Behind Taxi Drivers"

In partnership with Standard Bank CIB Digital and Mobalyz we are thrilled to bring you this hackathon in the exciting landscape of Internet of Things (IoT) data!

Taxis are an integral part of society in South Africa. They are critical to the movement of people and therefore operate in a way that is truly unique.

Vehicle telematics uses a combination of sensors, GPS systems, and vehicle diagnostics to collect and transmit vehicle data to a remote location. With telematics we can track the speeds, accelerations, location, and activity of vehicles. The sheer size of telematics data makes the process of understanding it even more complex and valuable.

Understanding the taxi driver is a complex task, and there are different types of styles and behaviours which they exemplify. We can't define a good or bad taxi driver by the same metrics we would use for the average road user. When faced with a fleet of taxis and drivers, there is true business value in identifying their key behavioural traits.

Behavioural segmentation is a common topic when faced with understanding people and their actions. It can take the form of unsupervised or supervised machine learning to classify, group or profile people into interpretable categories/profiles that accurately describe their behavioural patterns.

Using these behavioural profiles, we can start to understand and quantify risk profiles of our drivers. We can also start to answer questions around what makes a taxi driver 'good' or 'bad' and what type of drivers are suited to what type of environments.

Welcome to the Driver Behavioural Segmentation Challenge!

The hackathon aims to explore the vast potential of using telematics data to create comprehensive behavioural profiles of taxi drivers.

The form of your final output is not prescriptive, and creativity is encouraged. However, any categorisation of drivers must be interpretable.

When building your solution consider the following:

1. What other publicly available datasets can you find to augment your analyses?
2. What features are relevant when trying to characterise driver behaviour?
3. How does the geographical location affect driver behaviour?
4. How does continuous time spent driving affect drivers?

Bonus question (Not-compulsory):

Sometimes it is possible for there to be more than one driver associated to a particular vehicle id. However, this is not always explicitly stated and it's important to know when a particular vehicle is not being driven by the appropriate person. Can you identify any vehicles where there are multiple drivers?

Data

The dataset contains a sample of 106 unique vehicles based in the Western Cape. The dataset has been partitioned by vehicle id and is available as 106 separate csv files. A separate data dictionary will be provided.

An additional dataset including recorded insurance claims relating to accidents per vehicle is provided in the file named "claims_data.csv". These claims are related to a vehicle id and not necessarily the driver of the vehicle. They cover the period of January 2021 to January 2023. Dates of individual claims have been purposefully omitted and will not be provided.

Data will be provided to all entrants who have registered for the event via this SharePoint site. Soon after the launch event, you will be granted access to the Microsoft SharePoint site where the data and necessary files are provided.

Hints / Golden Nuggets

1. Please pay particular attention to dates and times when the driver is available (physically driving or active on the road)
2. Not all drivers have the same routes – consider different routes have different challenges (e.g. traffic lights, congestion, township routes)
3. Pay particular attention to the claims data and how you can augment it to create powerful insights
4. You are allowed to take multiple roles/views – it would be best to take one role (owner vs driver – to guide the solutions)
5. Try to emphasise specific actions taken from solutions you create (e.g. if I am a taxi owner – I'd like to understand the likelihood of my taxi being involved in an accident given driver behaviour or would I hire this driver? e.g. from a driver perspective: "how can I use IoT data to improve my driving skills?")

Team details

Teams can consist of 1 - 5 individuals. Approximately 6 teams will be shortlisted and then invited to present to a judging panel on Wednesday 23 August, 17:00 at an in-person event at the Stellenbosch University Museum. We fully encourage women in data science and there will be additional prizes for eligible teams.

Prizes:

Prizes will be award to the top three teams selected at the finals. The top-placed women's team will be eligible for an additional prize. The prizes are as follows:

1st Place Team – R22 000.00

2nd Place Team – R12 000 .00

3rd Place Team – R6 000.00

Women's Prize – R6 000.00

Please note to be eligible for the "Women's Prize" your team must meet one of the following conditions:

1. A team of 5 must have at least 4 female members.
2. In the case of a team with fewer than five members, all members must be female.

Submission and Judging:

The Hackathon will launch virtually on Thursday 10 August (18:30). Submissions will be accepted from Friday 18 August until Sunday 20 August (23:59)

Your submission should include the following:

- A pdf document of no longer than one page explaining your solution.
- A PowerPoint presentation describing the problem, your approach, and your solution. Finalists will be given an opportunity to polish their slides before the final presentations, without adding any content.
- A link to code. It can be a GitHub repository or a google drive link. For the latter, make sure that permissions are set so that we can download your code.
- All submitted documents must include your Team Name and Team Members
- Only one submission is permitted per team.

Make use of the submission form to submit your solution. The form will close on Sunday 20 August at 23:59.

There will be two rounds of judging. In Round 1, solutions will be scored and approximately 6 finalists will be selected to present their solutions at an in-person event on Wednesday 23 August. Finalists will be emailed on the evening of Tuesday 22 August; however, all participants are invited to attend the in-person event.

Industry professionals from Standard Bank CIB Digital and Mobalyz will be present at the finals, and this is a great networking opportunity.

Evaluation of solutions

Judges:

Round 1 Judges: Sunday Oladejo, Prof. Kanshu Rajaratnam, John Mukomberanwa, Nkosinathi Mntaka

Round 2 Judges: Claire Davis-Reddy (PhD), Martine Burgess, Dr. Sunday Oladejo, John Mukomberanwa, C.J. Naidoo, Pieta Heyns, GJ Cowper

Evaluation Criteria:

Category	Description
Problem interpretation (5)	Is the problem interpreted in a suitable manner? Is data used to drive modelling decisions?
Model design (10)	Are model design decisions well motivated?
Evaluation protocol (10)	Are metrics and evaluation procedures appropriately chosen?
Creativity and insights (10)	How creative is the solution approach? What ideas are presented as an extension to the base problem?
Interpretable Results (10)	Are the behavioural profiles easily interpretable and understandable for business users.
Relevance of solution to a context (12)	What impact could the solution have for Mobalyz?
Presentation and pitch (12)	Clarity and structure of the presentation.