Why is this a Problem?

Presumably Victoria always has some levels of accidents. Why is this a problem now for the Victoria Association of Local Government (VALG)? Why is it important? While no one wants people to die or get injured on the road, there is a limited budget in total and a lot of potential projects, so why would they pay money to PWC to analyse and reduce it? The VALG has to do road maintenance, waste disposal, etc., so that's a lot of tasks.

What metrics do we want to influence from this analysis?

Primary metric (see Appendix C)

- Road accident rate – calculation? I will take this as the total number of accidents over the 2-year period.

Secondary metrics:

- Rate of injuries – calculation? Why is 1 injury considered to be the same even if there are different levels of severity? Maybe I am free to create a metric here?
- Rate of deaths – calculation? I will take this as the total number of fatalities over the 2-year period

Why is the primary metric road accident rate and not rate of injuries or rate of deaths? Why the focus on road accident rate? While this is a testable hypothesis of amount of accidents correlation to injuries and deaths and it seems like it would be a strong correlation. Would it not be better to focus on the end result and then go backwards rather than state the primary metric and then hypothesis a causal relationship. Is there domain knowledge that states that road accident rate is more important? Or is it because of road maintenance costs caused by clean up post road accidents adds a second incentive to the VALG?

A rate implies a ratio of one item to another (i.e the rate of deaths per month or injuries per thousand kilometres drove), not an absolute amount. This is an important distinction as it is possible for a rate to fall while the total amount increases.

What is the deliverable required?

Report? Says it can be either a bunch of visualisations or a statistical analysis. Email mentions slide/documents/files/visualisations/supporting materials. Presumably there is a possibility that this may need to be presented based on the word "slides". This may mean that the analysis needs to be targeted towards that purpose.

What constraints are there on this Analysis

3 days before and after work. So cannot go to into depth on this, can't do a 6-pager detailing the methodology in total(see Appendix B) or prototype ideas or do too much external reading to improve domain knowledge or look at other people's analysis.

Can we integrate external data, or can we only use the data given?

Data – 2 years 2011 and 2012. Is this all the crash statistics or is it some? Seems to be mostly every crash one after the other but is it missing some crashes? The fact that the accident number is not in the same order as the date and time order suggests accidents have been entered late? Why would they be entered late? Are some accidents potential missing? Should I take it as all crashes or is it a biased sample? Can I validate the data? A lot of these questions possibly are not answerable in the timeframe.

To validate the data, I looked at the Victorian Road Safety Trauma Report to see if the numbers reconciled, they did on fatalities but not on the serious injuries (see Appendix A).

Appendix

<u>Appendix A</u>

	Year	2011	2012	Grand Total
DataSet	Serious Injuries	5075	4724	9799
	Deaths	283	280	563
Victorian Road Safety Trauma Report	Serious Injuries	5100	5400	10500
	Deaths	287	282	569
Serious Injuries	Serious Injuries	-25	-676	-701
	Deaths	-4	-2	-6
Serious Injuries	Serious Injuries	0%	-14%	-7%
	Deaths	-1%	-1%	-1%

Victorian Road Safety Trauma Report from below

https://www.vicroads.vic.gov.au/-/media/files/documents/safety-and-road-rules/victorianroadsafetytrauma2015.ashx

<u>Appendix B</u>

A 6-pager to me is a document which basically works as a 1-pager (i.e. quick summary) for the first page, then goes into more detail into return on investment, prototype design, analysis ideas, project plans, and final implementation methods.

For more information feel free to read these links

https://eugeneyan.com/writing/ml-design-docs/

https://writingcooperative.com/the-anatomy-of-an-amazon-6-pager-fc79f31a41c9

A 1-pager is a document which gives basic information on why you are doing the project, what you hope to get from it, and if there are constraints on the project. Quick document to keep people aligned in a project or to keep yourself on track. While a 1-pager which can be read about here.

https://eugeneyan.com/writing/what-i-do-before-a-data-science-project-to-ensure-success/#break-it-down-to-spot-rabbit-holes-and-dead-ends

I think good ideas set you up for good results and a great way of improving your ideas is to write them down.

<u>Appendix C</u>

A primary metric, in this document, is defined as the metric that the VALG has requested an improvement upon. The secondary metric is the item that they hope to influence by influencing the primary metric. This is important as changes to the secondary metric but that aren't correlated to the primary metric would imply that either the activities the VALG is doing to improve the primary metric is not working or that other items are stronger influencers.