Project Proposal for Petrol Pricing Prototype

Michael Spear - 2022

1 Executive Summary

This document aims to outline a Data Science prototype project which has the goal of proactive petrol cost minimisation. It proposes to achieve this by building a data pipeline to current petrol station prices, then do an exploratory data analysisof historical prices, before building a few basic cost minimisation models, and then measure the performance of those models versus a baseline of randomly simulated invididuals. The output will be 2 dashboards; one which will allow both operational users to make more informed decisions and another for management to see how much potential benefit is yet to be gained.

2 Motivation

Since 2010, Queensland uses on average 1,700 megaliters of petrol per month (A) at a price of between $1.30 to $.1.60 per liter (B), or $2.5 billion a month. In March 2022, the price of petrol passed $2 per liter (C), this inflationary pressure has impacted household living standards (D) and reduced corportate profits (E). If an application was made that could reduce the costs even by a few percent by understanding sources of variations, it could potentially produce quite a signficant amount of cost savings for individuals and organisations.

In an effort to improve market efficienes, the Queensland Government has mandated that all petrol stations report costs within 30 minutes of a price change (F). This increased level of pricing transparency makes this idea practical and gives a large amount of avaliable data.

3 Return on Investment

The proposed prototype is intended to give the following benefits:

- **5% reduced cost on weekly petrol cost versus baseline**. This is the key KPI for success of this project
- A dollar value of total controllable yearly petrol costs

The proposed costs are:

- $4,000 in labour costs (40 hours x $100/hr)
- $0 in software costs and hardware costs. Some software and hardware would incuur costs if moved into production
- $x in training costs from holding meetings explaining results

4 Requirements and Constraints

The main proposed outcomes of this project and their proposed benefits are:

1. **Create a report which shows controllable and uncontrollable petrol costs**. If all petrol prices are uncontrollable and therefore cannot be minimised, a company may wish to move their fleet to electric to minimise exposure or build storage facilities to capture petrol when it is cheap. If there is a high-level of controllability(i.e. certain petrol stations are always cheaper), companies/individiuals may make a series of heurestics to minimse cost or justidy more expensive machine learning algorithms and hardware.
2. **Create a dashboard which assists drivers minimise petrol cost by showing optimal destinations.** A cost savings of at least 5% over the baseline result is acheived.
3. **Create a monitoring dashboard to see how much of the cost minimisation benefit has been captured and how much is still on the table.** This will assist in reaching the 5% savings over baseline.

4.1 What's in-scope and out-of-scope?

Some problems are too big to solve all at once. Be clear about what's out of scope.

5 Methodology

5.1 Problem Statement

How will you frame the problem?

5.2 Data

What data will you use to train your model? What input data is needed during serving?

5.3 Datawarehouse Design

Creating facts and dimension tables based on star schema

5.4 Exploratatory Data Analysis and KPI Identification

Finding KPIs by using exploratory data analysis

5.5 Model Building

From the EDA build a simple model using heruistics or simple models

5.6 Simulation and Measurement Methodology

Creating a simulation from the data of 10000 people a year buying fuel from random locations, which is cheaper going to the (cheapest in surrounding suburbs vs model vs average in suburb) x average liters + average weekly budget (based on prior year simulations)

5.7 Business Intelligence

An image of a rough dashboard and how it can be used

6 Implementation

6.1 High-level Design

Start by providing a big-picture view.

6.2 Equipment

Using SQL, Python, API, Azure SQL Server, Power BI

6.3 Project Plan

Timeboxing is the main method. Want to complete in 1 week, as this is just to show that I do understand these key ideas.

A Appendix: Source of Australian Petrol Useage Data

See "Australian Petroleum Statistics - Data Extract January 2022" excel file and "Sales of petroleum products by State/Territory" tab. Sourced from https://www.energy.gov.au/publications/australian-petroleum-statistics-2022

B Appendix: Source of Australian Petrol Useage Data

Sourced from https://www.accc.gov.au/media-release/australian-petrol-prices-in-2020-21-were-lowest-in-22-years

C Appendix: March 22 Petrol Price Data

Sourced from https://www.msn.com/en-au/money/markets/fuel-price-relief-imminent-as-discount-phase-begins-in-parts-of-queensland/ar-AAVlBG0

D Appendix: Cost of Living Pressures

Sourced from https://www.abc.net.au/news/2022-03-15/petrol-prices-and-cost-of-living-jobs/100908296

E Appendix: Business Costs Rising

Sourced from https://www.northqueenslandregister.com.au/story/7658701/birdsville-bucks-the-fuel-price-trend-for-now/?cs=4770

F Appendix: Queensland Government rule on Petrol Price Changes

Sourced from https://www.fuelpricesqld.com.au/