

Data Science Skills Challenge

Document Version: 1.0

Date: 17/05/2022

Today is December 01, 2020. You work for **Vac-Attack** which sells **High-End Vacuums**. You are asked by the General Manager of Sales to use the historical data to create a Forecast of sales expected December 2020 as well as the total expected for the month. This will help them to determine whether the company will meet it's targets and ensures the stock on hand matches demand. You are also asked to derive any insights from your model or the data; especially around advertising spend.

Vac-Attack is primarily an As-Seen-On-TV business and therefore the majority of advertising is TV based. The Marketing team has said that the Marketing Mix for advertising has changed a little as the business has gone more digital. The Marketing team also believe the release of the Ultra Edition Vac was received positively and generated more sales.

Vac-Attack sells through an 0508 number and their website.

Data Definitions

The CSV is a historical daily record of sales and expected predictors that you have gathered from 2016 through to November 2020.

Date	Date
Positive News	A Positive News Story on Vac-Attack was released
Negative News	A Negative News Story on Vac-Attack was released
Competition	Vac-Attack runs competitions where you win a free Vac for your friend
AdvertisingSpend	Total Advertising Spend
Month	<i>Derived from Date</i>
Day	<i>Derived from Date</i>
0508 Line	A binary variable indicating when Vac-Attack had a 24/7 call line. They have since switched to fixed hours.
UltraEdition	A binary variable indicating when the Ultra Edition vac was made available.
Covid Lockdown Indicator	A binary variable indicating if lockdown was imposed or not
Sales	The units sold

In addition to the sales information above, the marketing team have provided the advertising buys (total spend) for December 2020 with no expectation of competitions or

news stories. The Ultra Vac will continue to be sold. The company aims to sell 3,900 units in December.

Files:

All the related data files are saved under the Data folder within the Challenge_Data.zip

- MarketingSales.csv – contains the data from marketing
- MarketingCols.csv – contains the column names for the marketing sales data
- DecemberAdData.csv – contains the advertising spend information for the month of December
- DecemberCols.csv – contains the column names for the DecemberAdData table.

How you will be assessed:

- You will be assessed based on how you understood the question and objectives, ability to explain and justify your approach, and your skill to communicate the results and methods to a non-technical audience.
- We also expect you to provide an explanation on why you decided to go for the approach you selected. What alternative methods are possible and what are the pros and cons of each alternative techniques?
- Lastly, your coding ability will be evaluated. We expect the code to be readable, reproducible, and easy to maintain/debug/understood by a reviewer.

Submission:

- Please submit your output in a format that we can read and open, either as a .py, .ipynb, .Rmd, .R, pdf, html, markdown file, or word.
- A re-runable code should be submitted as well -- ideally using R or python, whichever is your preferred language.
- Do not send any dataset. We would use the original dataset provided to run the code, so ensure that the codes are compatible to ingest the data in its original format.