

**Welcome to the Carvana Analytics experiment evaluation assignment!**

## Introduction

This project includes fictional data from an outbound call initiative Carvana has been making to users in order to attempt to increase sales.

In this exercise you will be asked to describe the results of how the campaign has been performing, explore how you might improve it, and evaluate a model provided by another member of the team.

We have set up this test in [Google Colab](#) where you can use SQL or python to complete it, but you may use any data analysis environment you wish to complete the tasks.

Items to return to us:

1. Answers to the following questions; both written and visualization for each in .pdf format.
2. A complete copy of the code you used in your analysis

The estimated time for this exercise is around 2 hours. Please submit your answers to your recruiting coordinator. Good luck!

## Data

- A complete data dictionary is located in Data Dictionary.xls
- The set of users can be found in UserHistory.csv
- The set of calls attempted can be found in OBCallHistory.csv
- The set of sales completed can be found in SaleHistory.csv
- Model Scores (for question 7) can be found in ModelScores.csv

*Individual users may have more than one call or sale attributed to them*

## Key Definitions

*Contact Rate* - the number of calls which resulted in customer contact divided by the number of calls made

*Sales Conversion Rate* - The number of users with a sale divided by the total number of users

## Questions

1. Does contact rate vary across time of day and/or day of week?
2. Does contact rate vary by the customer's age?
3. Does the sales conversion rate differ by Income (group this into \$5000 ranges)? Why might this be the case?
4. How does conversion rate vary by CreditScore2 (in 50-point buckets) with regard to contact status (contacted, attempted, not attempted)?
5. Has the sales conversion rate for contacted customers changed over time?

Do the following, then compose an email to stakeholders and executive leadership outlining your findings. Be sure to include supporting data and visualizations.

6. Design a simple, rule-based strategy for calling users and evaluate it versus the existing call data. Why did you choose this strategy? Do you think it will improve sales? If so, by how much? What proportion of users should we attempt to call?
7. A Data Scientist on the team created a model to improve the efficiency of calling. Each user's score is located in `modelScores.csv`. Using these scores, compare calling the top 25% of scores versus a strategy where we call 25% of your rule-based strategy from (6). Does the model improve sales? By how much? According to the model, what portion of users should we call?