

Frequently Asked Questions

This document contains all questions and answers from the Q&A calls on December 7th and January 5th as well as all questions sent to us by email. This document will continue to be updated over time as new questions come in and are answered.

Registration and Submission Logistics

Q: Can I register and then add a teammate later?

A: You can! If it won't bring your team to more than 4 people, you can email analyticsshowcase@carmax.com to add your new teammate.

Q: Can you give a better overview of the video segment? What are you looking for in terms of results and the report?

A: Video can include you speaking to results, showing graphs, or whatever you think will be helpful to synthesize the analysis you've done. You can find the full requirements for the video in our official rules at <https://analyticsshowcase.carmax.com/rules>.

Q: Should the report only present the findings? Can I include graphs and bullet points in my report?

A: The report is meant to supplement your submission video and should outline your work, methodology, and results. It can include graphs, bullet points, or other elements to support communicating your process and findings. You can find the full requirements for the report in our official rules at: <https://analyticsshowcase.carmax.com/rules>.

Q: Do I have to show myself in the video?

A: No, that is not required.

Q: Can I present in whatever format I like in the video?

A: Provided it meets our submission criteria, present your findings however you feel is appropriate to get your message across most effectively. We are primarily judging submissions on the accurate and insightful findings that are clearly communicated.

Q: How big is the competition?

A: We've done some campaigns across social media and other reach out campaigns, so we are hopeful to see people from all different backgrounds and schools across the nation.

Q: Who is the audience of the report, are they technical managers or business operations managers? **(Updated 1/6)**

A: Managers and leadership across analytics will be reviewing submissions, so you can expect them to have technical expertise and business analytics backgrounds.

Analysis

Q: Should I explain my methodology for how I analyzed the data and what insights I found?

A: The problem statement is primarily asking for your recommendation to our question, so you need to balance presenting your methodology in the report and video with the presentation of your findings. One category of judging is rigor – are your analyses sound and accurate? Demonstrating methodology is important for that.

Q: How should I analyze this data? What are you looking for?

A: We don't want to be too prescriptive about what to do with your analysis. There is not a single right answer to the problem statement – we are looking for you to give us your answer and support that answer. Try not to get too hung up on trying to anticipate what you think we want – do your best within the guidelines of the contest and find accurate insights within the data.

Q: What variables should I evaluate in my analysis?

A: There are multiple ways to look at analyzing the data. You can also derive other variables from what we've provided.

Q: Is there a specific software/programming language I should use for the showcase analysis?

A: No! You can use whatever app, software, or programming language you are familiar with to perform your analysis.

Q: Regarding the curiosity judging category, what does 'unique insights' mean?

A: When you look at the data you might find more 'obvious' insights that are very clear or apparent trends in the data. You will get a higher score in the curiosity category if you find some things that are deeper – make connections and find patterns in the data that go a bit beyond the surface.

Q: What kind of recommendations to CarMax are you looking for? Should it be sales recommendations? **(Updated 1/6)**

A: Not necessarily – we want you to look at the data and your insights to make any recommendations to answer the contest question about improving customer recommendations. These can come in a lot of different formats. There's no one right answer.

Q: I am curious about the definition of personalized shopping experience in CarMax – what current metrics does CarMax have to measure it? **(Updated 1/6)**

A: At CarMax personalization means to tailor a customer's shopping experience by using the information we know about them. We are looking for insights based on patterns in the data set that you think CarMax could act on to create a more personalized experience for CarMax shoppers.

Q: The challenge goal says, "how can CarMax use this information to improve business operations?" What types of associate support does CarMax currently have in a customer journey? **(Updated 1/6)**

A: We have customer experience centers where associates can help shoppers through chat or on the phone during their journey, as well as the associates who work in our stores where we sell cars. These existing touchpoints can be a recommendation point for change or improvement, but you can also go outside what currently exists with your recommendations.

Data

Q: I see there are two .zip files on GitHub. It looks like they are copies of the same data, is that correct?

A: There is only one data file on GitHub with 200,000 records, with a separate file for the data dictionary.

Q: Does the data solely include information about the vehicles or is there customer demographic information included/anything additional?

A: The data is mostly related to the vehicles, but also includes information about the market where the transaction took place. There is no specific customer data.

Q: Are we able to do research or should we stick with the data when running the analysis?

A: Key findings and your core analysis should come from the data. We're open to using outside sources to support your findings if they are properly cited. If most of your findings are coming from external data, you may want to reevaluate, and consider how the external research will apply considering the main competition dataset is masked.

Q: Is the masking random, or are they mapped over? If there's a stickiness to the make/model, are they all mapped the same way?

A: Every make and model will have a unique masking that does stick for the appraised and purchased vehicle information.

Q: Some values are variables instead of values – can I get more information about what they represent? Can I get another data point for the dataset?

A: We have to mask the data so we can share it publicly to protect proprietary information. We aren't able to provide additional insight into what those masked variables map to specifically, but you can still draw conclusions about variables using their masked acronyms (e.g. "I see this trend in purchases/appraisals with X attribute").

Q: Why are there null values?

A: This is real data, and if something shows up as null it means the data wasn't available. How you handle nulls in the data is up to you, but the null just means that the data doesn't exist – it does not mean that the value is 0.

Q: What does appraisal mean?

A: An appraisal is a condition assessment, coupled with a monetary offer, to purchase a vehicle from a CarMax customer.

Q: Does the online appraisal flag mean the car was also purchased in-store, or does it just mean the appraisal happened in person? **(Updated 1/6)**

A: The online appraisal flag is only for the appraisal and doesn't correspond with the type of store purchase.

Q: How should I treat outliers in the data (like people the trade in \$75k cars for \$5k cars)? **(Updated 1/6)**

A: You can use any data processing or statistical techniques in your analysis, so how you treat these outliers is up to you.

Q: Do the model numbers mean the same thing across makes? For example, are A_1 and B_1 similar things? **(Updated 1/6)**

A: Each make can have multiple models which align to the numbers in the model column, however each make had a unique process for designating these values, so across makes the model numbers aren't comparable.

Q: Is the "market" data referring to the geographical location of the market where the transaction took place? **(Updated 1/6)**

A: Yes, it's referring to the market where the transaction took place. CarMax assigns specific market areas and we have masked those in the data set, but they do map to geographic locations nationally.

Q: Why do columns like mpg, fuel capacity have so few unique values? Should i think of it as a numerical column with discrete values or categorical? **(Updated 1/6)**

A: This is because most of these are values assigned by car manufacturers, and cars are made with similar specifications and capabilities. You will find a lot of the variables are not just

numeric, but can be treated as categorical as well. Use your best judgment for what you think will work for your analysis.

Q: Did all of these transactions occur in 2022? **(Updated 1/6)**

A: Not necessarily, this was a random sample of data so could have happened in 2022 or earlier.

Q: Is it encouraged to pre-process the dataset before carrying our analysis? For instance, should I remove all the rows containing "NA" values? **(Updated 1/6)**

A: We want to make this an open space for you to utilize your creativity to make decisions that you think will be helpful. If removing the data makes your analysis more meaningful, you can do so. But you are also able to run the analysis without removing the data - you should use your best judgment to find the best solution possible.

Q: Is one market more important than another? Can we assume that the markets are similar based on their density or sales performance? **(Updated 1/6)**

A: The data set includes a sample of markets from across the US and each market can include more than one store, so some may have higher volume than others. Because of the sample, each market might not be alike in the data set.

Final Round

Q: What happens if my team is selected as a finalist?

A: Finalists will be notified on January 22nd and present virtually in the afternoon of February 17th and will be released no later than 4:00PM ET. Ahead of the presentation, finalists will be partnered with a coach to run through things and help finalize the presentation. At that point, all the data and work will have been set. We will make sure finalists have the resources they need well ahead of the February 17th presentation day.