

Case Study Project – Video Game Categories

This document provides all the details you need to complete the case study. We recommend you read this document carefully before you get started.

Company Background

Snow Cone Games are a digital retailer for video games. The website includes a simple filter for the genre of the game as well as more advanced filtering. The product manager has recently learned that only 23% of buyers search for games using the genre filter.

71% of buyers use advanced filters to find the types of games that appeal to them. The product manager hopes that new categories based on different game attributes can help players find what they are looking for faster.

Customer Question

The product manager would like to know:

- Can you use the data from 40,000 games to identify groupings of games that are most similar to each other?

Success Criteria

Ideally, there would be no more than five groupings. Any more would make it much harder to display on the page so there would need to be a good reason to have more.

Data Dictionary

The data is available in a DataCamp Workspace, which you can find from the certification dashboard. The data set has the following columns:

Column Name	Details
name	Character, name of the game title.
release_date	Character, release day of the video game.
developer	Character, name of the developer that created the video game.
achievement	Numeric, number of 'achievements' players can earn in the game.
original_price	Character, the price of the video game (or "Free" for free titles).
percent_positive	Character, the percentage of players who review the game positively.
num_reviews	Numeric, the total number of players who reviewed the game.
multiplayer	Character, whether or not the game has multiplayer (True) or not (False).
mature	Character, whether or not the game has mature content unsuitable for younger players (True) or not (False).

Submission Requirements

1. You are going to create a written report to summarize the analysis you have performed and your findings. This report should be for the data science manager. The task list below describes what they expect to see in your report.
2. You will need to use DataCamp Workspace to complete your analysis, write up your findings and share visualizations.
3. You must use the data we provide for the analysis.
4. You will also need to prepare and deliver an oral presentation. You should prepare around 8-10 slides to present to the non-technical customer. The task list below describes what they expect to see in the presentation.
5. Your presentation must be no longer than 10 minutes.

Task List- Written Report

Your written report should include both code, output and written text summaries of the following:

- Data validation, including a summary of any changes you make to the data
- Exploratory Analysis, including graphics to support your findings
- Model Development, including justification for your choice of models
- Model Evaluation, including explanation of what this means about your models
- Comparison to the business success criteria
- Final summary including recommendations for future work that the business should undertake

Task List - Oral Presentation

Your presentation should be targetted at the non-technical customer who requested the work you have completed. The presentation should include:

- An overview of the project and business goals
- A summary of the work you undertook and how this addresses the problem
- Your key findings including how your work compared to the business success criteria
- Your recommendations to the business for future work

Grading

Before submitting your written report or delivering your oral presentation, remember to check your work against the following grading criteria. You must pass all criteria to pass this part of the certification.

Domain	Description	Sufficient	Insufficient
Data Validation	Assess data quality and perform validation tasks	Has validated all variables and where necessary has performed cleaning tasks to result in analysis-ready data.	Has not conducted all the required checks and/or has not cleaned the data. May have removed data rather than performed cleaning tasks.
Data Visualization	Create data visualizations in coding language to demonstrate the characteristics of data and represent relationships between features.	Has created at least two different types of data visualization that highlight characteristics of variables after validation. Has created at least one visualization that shows the relationship between	Has used the same visualization throughout. Has not included graphics to represent single variables and relationships. Has not used visualizations that

		<p>two variables.</p> <p>Has used visualizations that support the findings being presented.</p>	<p>support the findings being presented.</p>
Model Fitting	Implement standard modeling approaches for supervised or unsupervised learning problems	<p>Correctly identified the type of problem (regression, classification or clustering)</p> <p>Has selected and fitted a model for that problem to be used as a baseline.</p> <p>Has selected and fitted a comparison model for the problem that they were provided.</p>	<p>Has incorrectly identified the type of problem.</p> <p>Has not fitted a baseline model or has used a model for the wrong type of problem.</p> <p>Has not fitted a comparison model or has used a model for the wrong type of problem.</p>
Model Evaluation	Use suitable methods to assess the performance of a model	<p>Compared the performance of the two models/approaches using any method appropriate to the type of problem.</p> <p>Has described what the model comparison shows about the selected approaches.</p>	<p>Has selected a method not suitable for the type of problem.</p> <p>Has not described what the results show about the selected approaches.</p>
Business Focus	Make recommendations for analytic approaches based on business goals	<p>Has described at least one of the business goals of the project</p> <p>Has explained how their work has addressed the business problem</p> <p>Has provided at least one recommendation for future action to be taken based on the outcome of the work done</p>	<p>Has not identified any business goals</p> <p>Has not explained how their work has addressed the business problem</p> <p>Has not provided any recommendations for future actions</p>
Business Metrics	Judge performance of analytic results against relevant business criteria	<p>Has defined a KPI to compare model performance to business criteria in the problem</p>	<p>Has not identified a KPI to compare the model performance to the business problem</p>

		Has compared the performance of the two models/approaches using the defined KPI	Has not compared the performance of the two approaches using the defined KPI
Communication	Employs multiple tactics (written and verbal) to communicate to business leaders	<p>For each analysis step, has provided a written explanation of their findings and/or reasoning for selecting approaches</p> <p>Has delivered a verbal presentation addressing the business goals, outcomes and recommendations</p>	<p>Has not provided a written summary for each step</p> <p>Has not delivered a verbal presentation</p>