

# Analysis of Predictors of Price and Popularity of Steam Games in 2016

## Introduction

This is a dataset on all steam games scraped from public Steam APIs and steamspy.com. The dataset comes from data.world at this url: <https://data.world/craigkelly/steam-game-data>. The dataset has 78 columns and includes things like price, descriptions and release date. According to the data on Github, the repository was created in September 12, 2016 so we are assuming that the data also comes from the end of the year in 2016. The dataset has the following columns:

QueryID - (Integer) The original ID in idlist.csv

ResponseID - (Integer) The ID returned in the Steam response (should equal QueryID)

QueryName - (Text) The original name in idlist.csv

ResponseName - (Text) The name returned in the Steam response (should equal QueryName)

ReleaseDate - (Text) Appears to be the initial release date for the game

RequiredAge - (Integer) list named required\_age in JSON

DemoCount - (TextualCount) list named demos in JSON

DeveloperCount - (TextualCount) list named developers in JSON

DLCCount - (TextualCount) list named dlc in JSON

Metacritic - (Integer) numeric score from metacritic object in JSON

MovieCount - (TextualCount) list named movies in JSON (used object id for unique count)

PackageCount - (TextualCount) list named packages in JSON

RecommendationCount - (Integer) from recommendations.total in JSON

PublisherCount - (TextualCount) list named publishers in JSON

ScreenshotCount - (TextualCount) list named screenshots in JSON

AchievementCount - (Integer) achievements.total in JSON

AchievementHighlightedCount - (TextualCount) for achievements.highlighted in JSON

ControllerSupport - (Boolean) True if controller\_support was full

IsFree - (Boolean) is\_free in JSON

FreeVerAvail - (Boolean) True if is\_free\_license is True in package\_groups list

PurchaseAvail - (Boolean) True if price\_in\_cents\_with\_discount greater than 0 in package\_groups list

SubscriptionAvail - (Boolean) True if is\_recurring\_subscription is True in package\_groups

PlatformWindows - (Boolean) True if platforms.windows is True

PlatformLinux - (Boolean) True if platforms.linux is True

PlatformMac - (Boolean) True if platforms.mac is True

PCReqsHaveMin - (Boolean) True if pc\_requirements.minimum is non-empty string

PCReqsHaveRec - (Boolean) True if pc\_requirements.recommended is non-empty string

LinuxReqsHaveMin - (Boolean) True if linux\_requirements.minimum is non-empty string

LinuxReqsHaveRec - (Boolean) True if linux\_requirements.recommended is non-empty string

MacReqsHaveMin - (Boolean) True if mac\_requirements.minimum is non-empty string

MacReqsHaveRec - (Boolean) True if mac\_requirements.recommended is non-empty string

CategorySinglePlayer - (Boolean) True if for any i, categories[i].description is "single-player"

CategoryMultiplayer - (Boolean) True if for any i, categories[i].description is one of: "cross-platform multi-player", "local multi-player", "multi-player", "online multi-player", "shared/split screen"

CategoryCoop - (Boolean) True if for any i, categories[i].description is one of: "co-op", "local co-op", "online co-op"

CategoryMMO - (Boolean) True if for any i, categories[i].description is "mmo"

CategoryInAppPurchase - (Boolean) True if for any i, categories[i].description is "in-app purchases"

CategoryIncludeSrcSDK - (Boolean) True if for any i, categories[i].description is "includes source sdk"

CategoryIncludeLevelEditor - (Boolean) True if for any i, categories[i].description is "includes level editor"

CategoryVRSupport - (Boolean) True if for any i, categories[i].description is "vr support"

GenreIsNonGame - (Boolean) True if for any i, genres[i].description is one of: "utilities", "design & illustration", "animation & modeling", "software training", "education", "audio production", "video production", "web publishing", "photo editing", "accounting"

GenreIsIndie - (Boolean) True if for any i, genres[i].description is "indie"

GenreIsAction - (Boolean) True if for any i, genres[i].description is "action"

GenreIsAdventure - (Boolean) True if for any i, genres[i].description is "adventure"

GenreIsCasual - (Boolean) True if for any i, genres[i].description is "casual"

GenreIsStrategy - (Boolean) True if for any i, genres[i].description is "strategy"

GenreIsRPG - (Boolean) True if for any i, genres[i].description is "rpg"

GenreIsSimulation - (Boolean) True if for any i, genres[i].description is "simulation"

GenreIsEarlyAccess - (Boolean) True if for any i, genres[i].description is "early access"

GenreIsFreeToPlay - (Boolean) True if for any i, genres[i].description is "free to play"

GenreIsSports - (Boolean) True if for any i, genres[i].description is "sports"

GenreIsRacing - (Boolean) True if for any i, genres[i].description is "racing"

GenreIsMassivelyMultiplayer - (Boolean) True if for any i, genres[i].description is "massively multiplayer"

PriceCurrency - (Text) price\_overview.currency in JSON

PriceInitial - (Float) price\_overview.initial in JSON, divided by 100.0 to convert cents to currency

PriceFinal - (Float) price\_overview.final in JSON, divided by 100.0 to convert cents to currency

SteamSpyOwners - (steamspy.com) total owners, which includes free weekend trials and other possibly spurious numbers.

SteamSpyOwnersVariance - (steamspy.com) total owners, which includes free weekend trials and other possibly spurious numbers. Note that this is not technically variance: according to steamspy.com, "the real number... lies somewhere on... [value +/- variance]"

SteamSpyPlayersEstimate - (steamspy.com) best estimate of total number of people who have played the game since March 2009

SteamSpyPlayersVariance - (steamspy.com) errors bounds on SteamSpyPlayersEstimate. Note that this is not technically variance: according to steamspy.com, “the real number... lies somewhere on... [value +/- variance]”

SupportEmail - (Textual) support\_info.email in JSON

SupportURL - (Textual) support\_info.url in JSON

AboutText - (Textual) about\_the\_game in JSON

Background - (Textual) background in JSON

ShortDescrip - (Textual) short\_description in JSON

DetailedDescrip - (Textual) detailed\_description in JSON

DRMNotice - (Textual) drm\_notice in JSON

ExtUserAcctNotice - (Textual) ext\_user\_account\_notice in JSON

HeaderImage - (Textual) header\_image in JSON

LegalNotice - (Textual) legal\_notice in JSON

Reviews - (Textual) reviews in JSON

SupportedLanguages - (Textual) supported\_languages in JSON

Website - (Textual) website in JSON

PCMinReqsText - (Textual) text of pc\_requirements.minimum

PCRecReqsText - (Textual) text of pc\_requirements.recommended

LinuxMinReqsText - (Textual) text of linux\_requirements.minimum

LinuxRecReqsText - (Textual) text of linux\_requirements.recommended

MacMinReqsText - (Textual) text of mac\_requirements.minimum

MacRecReqsText - (Textual) text of mac\_requirements.recommended

Our goal is to discover interesting trends that allow us to predict how a game is price and it's popularity after release. The price is divided into price initial and price final. We are going to use price initial as our metric to determine price and SteamSpyOwners as a metric to predict popularity.

## Data Cleanup

From the above, we can see that there is a lot of information that is hard to use to predict price. One example is the description. It is hard to