Project-3: Capability Analysis & Project Proposal

Video Game Reserves

Team Pie

Data Analytics
Prof Karim
4/12/18

Table of Contents

Overview

The goal of this report to analyse a dataset (NCSA reserves) and identify how the data can be used to create an effective data product. Alongside the report also identifies potential challenges in building the data product as well as its benefits.

Part A: Data Product

Dataset

The dataset chosen is “NCSA reserves”[1]. It consists of reserves (pre-orders and pre-purchases) data of more than 60 games recorded every week across different channels, regions and platforms. The data captured is cumulative reserves to date since game announcement. All reserves figures are in actual units. The dataset contains 31,704 rows and 7 columns. A complete list of the variables, their description and type is provided in the appendix.[Appendix-1]

```
C:\Users\abatra\PycharmProjects\trial\venv\Scripts\python.exe '
The original dataset has 31704 rows, and 8 columns
Unnamed: 0        int64
Edition          object
Channel          object
Region           object
Installment      object
Platform         object
RelativeWeek      int64
ReservesLevel     int64
dtype: object

Process finished with exit code 0
```

Capabilities/ Features Enabled

1. Trends Dashboard
    a. Strategic Planning Team
        i. Channel evolution over time: Allows users to see channel evolution across regions over time.
        ii. Platform evolution across regions: Allows the user to see the platform evolution across regions as well as identify early adopters and laggards for new consoles.

iii. Business Territory Share (Region evolution): Allows user to understand changing market landscape and thus identify high growth potential regions.
iv. Franchise games evolution (Follow up games): Allows users to understand target audience response to franchise games and their potential for sequel launches.

b. Marketing Team
i. Impact of changing user preferences: Identify channel and platform preference across regions and allocate marketing budget accordingly.

2. Predict Tool
   a. Reserves per region
   b. Reserves per platform
3. Other Decisions
   a. Retail Planning: The team can use the final reserves numbers to derive potential 1st week sales and thus plan for retail (physical) manufacturing and inventory planning.
   b. Customer Support: Utilize reserves predictions as indicators of 1st week sales and use it to staff the customer service center thereby helping improve customer service and brand image.

Benefits

User Perspective

1. Understand changing marketplace trends
2. Improved work experience: View analysis vs do analysis
3. Same Language: Consistent data across teams and geographies

Business Perspective

1. Direct Impact
   a. Identify opportunities (Reserves prediction per region and platform)
   b. Leverage identified opportunities (Marketing spend across regions)
2. Indirect Impact
   a. Customer success team size (incidents dependent on 1st week game sales which are derived from final reserves)
   b. Franchise sequel release decisions

Data Description

To develop the features mentioned above, it is important to identify the key attributes, their impact and their relationship. From this section forward, we will be considering only the trends dashboard and the predict tool since the additional functionalities can be easily derived from the first two capabilities mentioned.

Data Points

Data Point	Why is it important?
Platform	With time each platform evolves. Install base for new platforms increase thereby boosting potential buyers. Also each platform has its own curve and can greatly impact final reserves level.
Region	Each region has its own culture and thus the reserves of each game depends on multiple factors including region culture.
Channel	With time user preferences change in terms of physical vs digital. This is also partly impacted by technology present at the time. Each channel has its own curve and can greatly impact final reserves level.
Edition	Each edition comes is a separate offering and has a curve of its own.
Release Date	Release date play an important role. A release in Mar is received differently from a game launched in Oct.
Relative Week	The number of weeks available for users to reserve a game post announcement date impacts the final reserves and thus make relative a critical factor.
Reserve Level	Indicator of reserves for similar games which can be used to derive a curve.

Weights

Though weights are important to derive the results but deriving weights for each attribute is difficult without looking at underlying dataset in depth.

Connectedness

Trends Dashboard

The features included in the dashboard are simple visualizations of the historical data and the relationship between the required variables is almost non-existent and hence not mentioned here.

Predict Tool

Connecting data from multiple attributes like edition, channel, platform and region are indicators of user preference. This combined with relative week and reserves level indicate a curve unique to each channel, platform, region combination which is critical to derive an accurate reserves level and thus 1st week sales.

Part B: Data Integrity

Data Related Challenges

1. Incomplete data: Looking at the data, it is easy to see that data is incomplete. E.g.: Reserves level for last 2 days are available for only a few games.
2. Skewed data: Games with different launch quarters have different reserves curves. Since most games are launched in Oct-Nov timeframe, they are not good indicators for games launched in other months.

Validating Assumptions

1. Impact of relative week: As expected, reserves level increase as launch week gets closer.
2. Correlation Matrix
   Since only two of the variables are continuous numeric, the correlation shows how they are related. As the other independent variables are categorical, a regression analysis is required to understand how they impact the reserve level.

```
                RelativeWeek  ReservesLevel
RelativeWeek        1.000000      -0.050785
ReservesLevel      -0.050785       1.000000

Process finished with exit code 0
```

3. OLS Regression Results
   (See on next page)

```
                            OLS Regression Results
==============================================================================
Dep. Variable:          ReservesLevel   R-squared:                       0.110
Model:                            OLS   Adj. R-squared:                  0.110
Method:                 Least Squares   F-statistic:                     207.1
Date:                Thu, 12 Apr 2018   Prob (F-statistic):               0.00
Time:                        08:15:58   Log-Likelihood:            -3.6629e+05
No. Observations:               31704   AIC:                         7.326e+05
Df Residuals:                   31684   BIC:                         7.328e+05
Df Model:                          19
Covariance Type:            nonrobust
================================================================================
                   coef    std err          t      P>|t|      [0.025      0.975]
--------------------------------------------------------------------------------
Intercept     -1.72e+04   6429.186     -2.676      0.007   -2.98e+04   -4601.540
Region[T.CAN] -1689.0392    542.819     -3.112      0.002   -2752.986    -625.093
Region[T.MEX]  -861.9661    600.002     -1.437      0.151   -2037.993     314.061
Region[T.ROSA]  690.4120    631.862      1.093      0.275    -548.062    1928.886
Region[T.USA]  1.014e+04    541.080     18.741      0.000    9079.918    1.12e+04
Platform[T.PC] 8551.0669   5160.784      1.657      0.098   -1564.270    1.87e+04
Platform[T.PS VITA] 2707.5000 6133.860   0.441      0.659   -9315.105    1.47e+04
Platform[T.PS3] 1.935e+04   5182.742      3.734      0.000    9193.358    2.95e+04
Platform[T.PS4] 1.143e+04   5158.192      2.216      0.027    1320.004    2.15e+04
Platform[T.Switch] 3026.8896 5467.590     0.554      0.580   -7689.799    1.37e+04
Platform[T.Wii]  211.2497   5452.774      0.039      0.969   -1.05e+04    1.09e+04
Platform[T.Wii U] 4355.2008 5245.012      0.830      0.406   -5925.226    1.46e+04
Platform[T.XBOX ONE] 1.094e+04 5158.716   2.121      0.034     830.955    2.11e+04
Platform[T.XBOX360] 3.392e+04 5193.624    6.531      0.000    2.37e+04    4.41e+04
Channel[T.Retail] 3999.5967  360.305     11.101      0.000    3293.384    4705.809
Edition[T.Collector]  950.9742 3829.135   0.248      0.804   -6554.280    8456.228
Edition[T.Deluxe] 6660.8291 3831.316      1.739      0.082    -848.700    1.42e+04
Edition[T.Gold] 7292.0551 3818.271        1.910      0.056    -191.905    1.48e+04
Edition[T.Standard] 9888.1499 3810.720    2.595      0.009    2418.991    1.74e+04
RelativeWeek  -113.5991    7.335    -15.487      0.000    -127.976     -99.222
==============================================================================
Omnibus:                    58547.197   Durbin-Watson:                   0.186
Prob(Omnibus):                  0.000   Jarque-Bera (JB):         99326217.884
Skew:                          13.915   Prob(JB):                         0.00
Kurtosis:                     275.792   Cond. No.                     3.36e+03
==============================================================================

Warnings:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
[2] The condition number is large, 3.36e+03. This might indicate that there are
strong multicollinearity or other numerical problems.

Process finished with exit code 0
```

As can be seen because of limited reserves in some regions and for some platforms they have little impact on the regression as their P values are above 5%. These values will still be of importance in predicting future sales in those regions and for those platforms as game vendors will want to know where to spend time and effort to on a game.

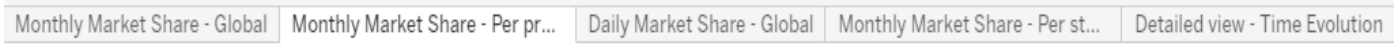
Feature Improvement

As trends change with every passing year, it is important to include all the variables (including categorical variables) to ensure that all dependencies are accounted for and an accurate prediction model is built.

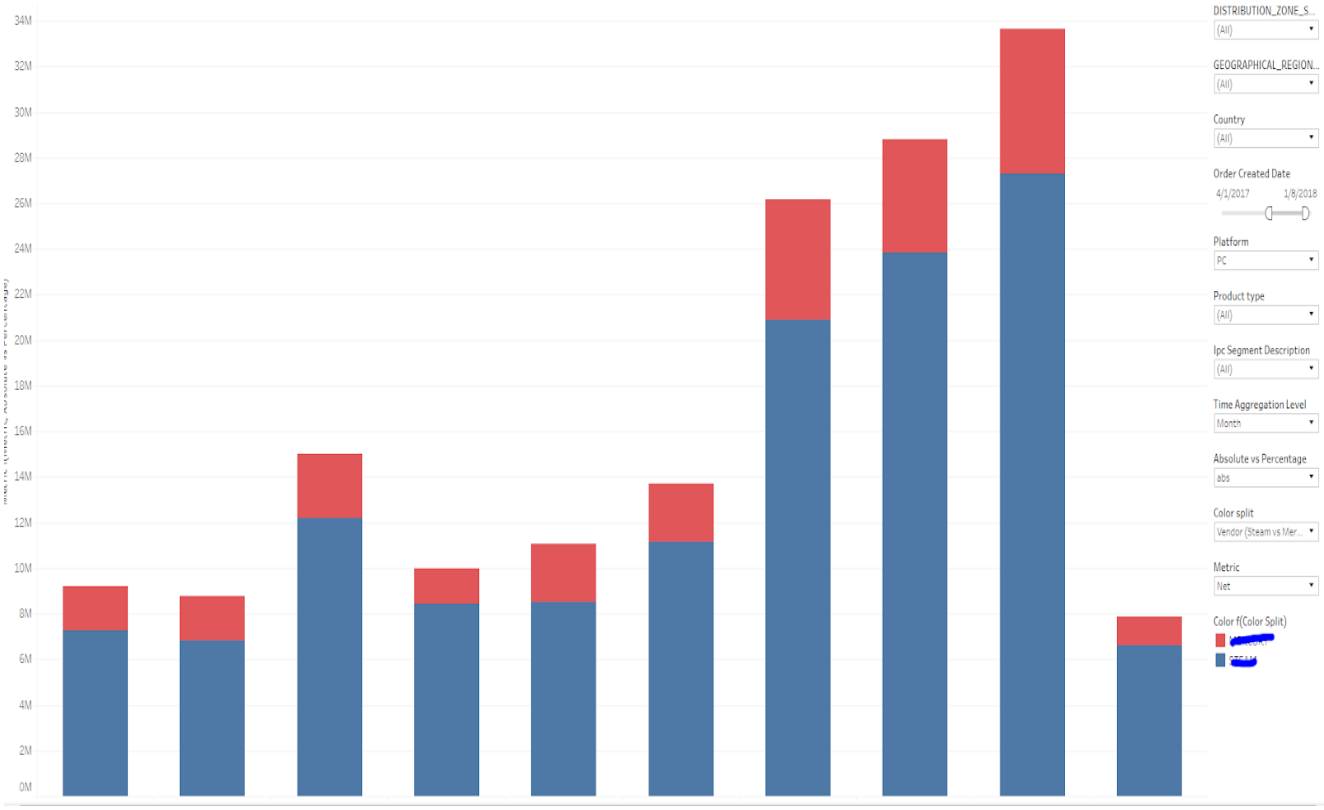
Part-C: User Representation

User Representation

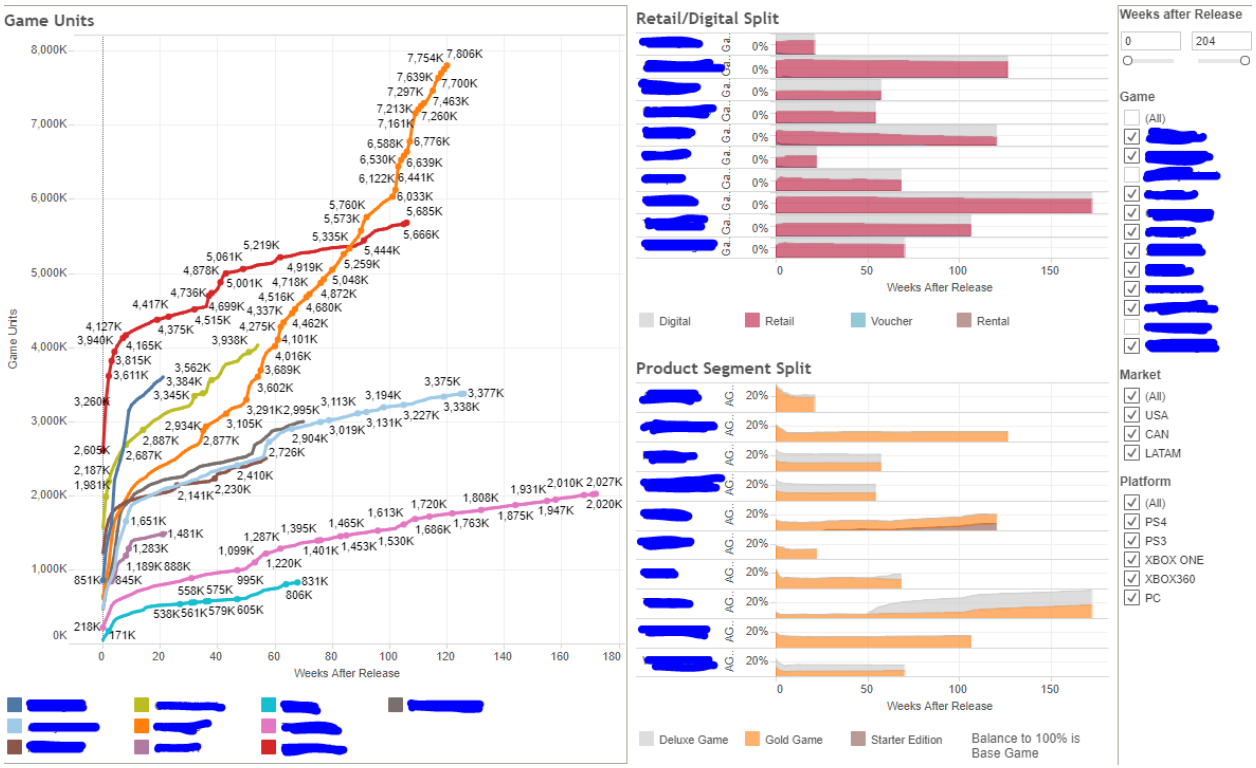
The product is presented to the user as an interactive dashboard (very similar to a Tableau dashboard). The dashboard will consist of multiple tabs which will ensure that the data is presented in a methodical fashion to the user. A simple example of this can be seen in the figure below:

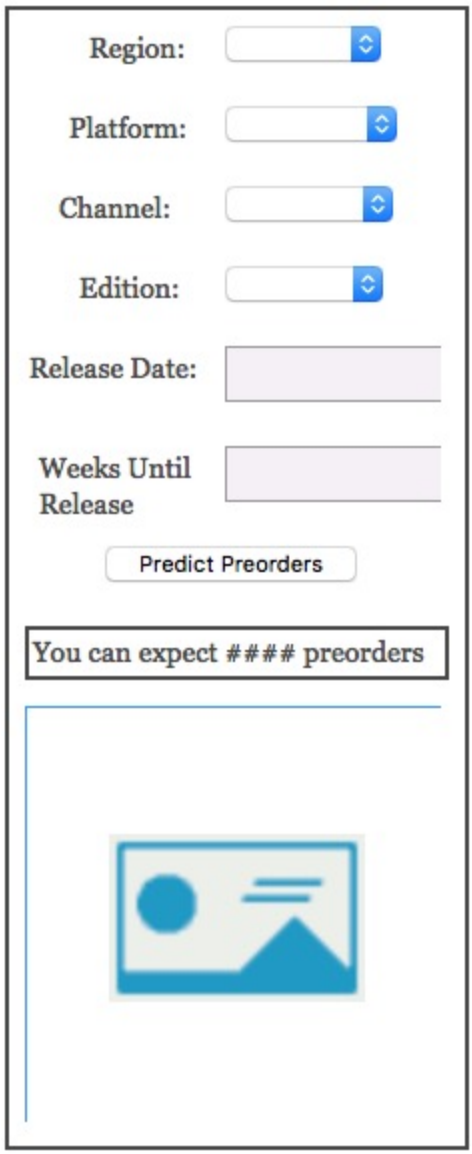


Within each tab, the user will have access to different trends and will be able to interact with the data live using filters.



The prediction tool will also have a dashboard like appearance which will predict reserves curve for a game based on release date, region, platform as well as comps (competitive titles) chosen. The dashboard will also showcase the reserves trend for the comps (competitive titles) chosen for reference.





Product Validation

Any data product is only as good as its effectiveness. To improve the product's effectiveness, it is important to continuously analyze the performance of the feature. Some key success metrics that can help evaluate the content discovery & personalization feature are:

	Metrics	Why?
Direct Indicators	Accurate reserves predictions	The key goal of the tool is to predict reserves. Inaccuracy can make the tool ineffective and thus invaluable.

	Decreasing time spent on forecasting	If a user is able to forecast new titles with ease without spending a lot of time then it indicates that the product is working well.
Indirect Indicators	Easier collaboration across teams	Single data view allows users across teams to be on the same page and thus collaborate more effectively with other teams.

Test & Autocorrect

The product can be set up such that it automatically pulls the digital reserves data from partner portals and check for accuracy. It can then add the data to its existing dataset to generate a new training set and thus an improved model

References

1. DA_2018_Homework-2_Batra, https://docs.google.com/document/d/1u8WvIfCbwrR51Yi_5kEhF_CqX4HZ51UkiQad8CQJqsA/edit#
2. DA_2018_Project-2_Batra, https://drive.google.com/open?id=1oy9ZeRNZ2VSm71MCXx4cw9PVBqc5U4-B3O51J3gzvGs

Appendix - I

Data Variables

Rows: 31,704
Columns: 7
Link: https://drive.google.com/open?id=1mXo7K4LCez55Re17QGqYaDcAsyWl_fkg

Column Metadata

Variable ID	Variable Descp
Edition	Game offering: Standard, Gold, Deluxe
Channel	Distribution Channel i.e. retail or digital
Region	NCSA specific further divided into US, CAN, MEX, BRA, ROSA
Installment	Game name
RelativeWeek	Relative week before game launch
ReservesLevel	Game reserves (pre orders + pre purchase)
Retailer	Retailers or business partners distributing the game.
Platform	Console for which the game is reserved
ReleaseDate	Game launch date