Future Sales Predictions on Russian Electronics Shops

Daniel Roberts, Daniel Simpson, Jacob Sweeten, Edward Mantsevich
Advisor: Dr. William Eberle
Department of Computer Science, Tennessee Technological University

Objective

Given a dataset of sales from a chain of Russian electronics shops from January 2013 - October 2015, generate sales predictions for the month of November 2015.

Introduction

The prediction of future sales is a pervasive problem across many industries. To contribute towards further solutions to this problem, 1C Company has provided a Russian sales dataset to the Kaggle competition, “Predict Future Sales.” The goal of the challenge is having participants predict the future sales of products for the next month extrapolated past the end of the dataset. Our team will seek to analyze the dataset and produce models that are able to generate reasonably accurate predictions using some subset of the data that is known. Using these analyses, we will have a metric against which to assess the model’s performance when predicting values outside the scope of the dataset.

Dataset

The dataset for this project was provided by the Russian firm 1C Company and contains sales data for a number of different products per day for a total of 60 store locations in Russia.

1C Company is a software developer, distributor and publisher, and as such primarily focuses on the sale of electronics and media. Hence, the products incorporated in this dataset include:

- copies of major game titles for multiple platforms, game consoles and their associated accessories
- merchandise related to game franchises
- educational and antivirus software packages
- computing accessories
- music discs and audiobooks.

This dataset spans the time period of January 1, 2014 through October 31st, 2015. The product names are recorded in Cyrillic script.

Research Questions

Given this sales dataset of Russian electronics shops, our major goal is predicting sales for the month of November 2015. We also seek to meet a number of smaller goals including:

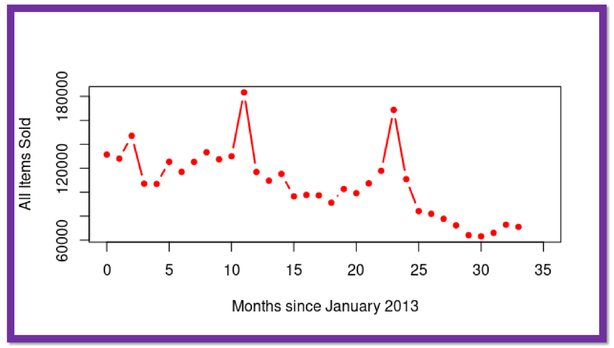
- Overcome the lingual barrier of working with a Russian dataset in Cyrillic script.
- Confirm that our prediction of observing large spikes in sales around Christmas in 2013 and 2014 is accurate.
- Discover if items such as the PS3 controllers or Kaspersky Antivirus licenses which we hypothesize will be more stable sellers are actually stable.
- Given the popularity of electronic items, games, and media, we hypothesize that the company in general will be fairly stable with reasonable sales across most, if not all, categories at many store locations. Since we have very little information about individual store locations, we will speak more generally about sales across all store locations.

Exploratory Analysis Methods

The majority of our analysis takes advantage of the statistical software environment R, making use of the packages *forecast*, *dplyr*, and *ggplot2*. We have performed a number of tests to gather observations about this dataset including:

- translating portions of it in order to better understand what an item is
- plotting sales against one another over time in order to visually observe trends in sales
- basic statistical analysis to observe possibly erroneous data or outliers and to better understand the scale of the sales in the dataset

For instance, negative values in sales could be interpreted as retail “shrink” (an industry term for decrease in profit margin due to a number of factors, including theft), or item returns. Ultimately, we are seeking to gather sufficient useful data to set up predictive models such as the Holt-Winters predictive model to generate sales predictions on this dataset.



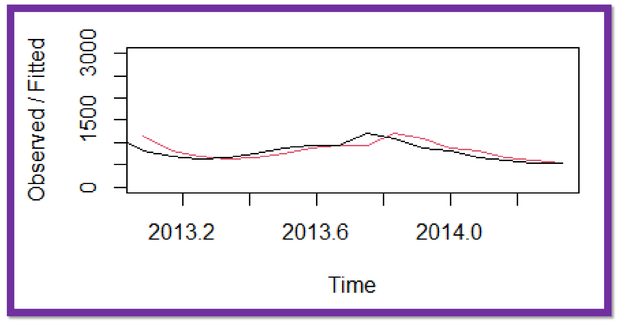
The above chart displays the sales plot for all items sold at all store locations for the full duration of the dataset.

Predictive Methods

- Initial goal was to use Holt Winters method to create predictions of the sales data of the following test items:
  - Diablo 3 PC Edition
  - PS3 DualShock Controller
  - Kaspersky Antivirus
- Unfortunately, these items do not seem to follow any sort of seasonal trend, so the predictions we obtained from Holt Winters were quite inaccurate.
- With this knowledge, we fell back on the Simple Exponential Smoothing method which is better suited for non-seasonal data.
- The Simple Exponential Smoothing involves a weighted average of the past data that assigns a higher weight to recent data. This means that the weight decreases exponentially as the age of the data increases.

Product Name	Jan 2014 Prediction	Jan 2014 Actual	Jan 2014 % Error	June 2014 Prediction	June 2014 Actual	June 2014 % Error
Diablo 3 PC	310	235	31.91	430	305	40.98
Kaspersky Antivirus	879	822	6.93	546	564	3.19
PS3 DualShock	282	301	6.31	282	190	48.42

The above table displays the results of two different Simple Exponential Smoothing predictions for the monthly sales of three test items



The above chart displays the forecast line of the Simple Exponential Smoothing algorithm over Kaspersky Antivirus sales from Jan 2013 through May 2014.

Findings and Observations

Given the nature of the data, we believe a time-series prediction method is best suited for our needs. We have found that Holt Winters Method was less accurate than Simple Exponential Smoothing Method when tested on individual items over the periods of data we have. However, the Holt Winters model may prove to be more useful in predicting total sales of all items over the entire period.

Looking at the prediction data from the given table, it can be observed that an item with high variance in its sales data, such as Diablo 3, is hard to predict accurately. This makes sense because video games experience drastic changes in popularity. The other two items produced more accurate results, but surprisingly the predictions for the DualShock controller were less accurate on average than that of Kaspersky Antivirus, despite the DualShock having more consistent sales. This prompts further queries into the correlation between sales consistency and prediction accuracy.

We had hypothesized that certain items such as Kaspersky Antivirus licenses would sell consistently, but it actually yielded less predictable sales. However, we found that Play Station controllers sold more consistently than expected. These findings demonstrate that we need to find a more reliable approach to predicting the consistency of item sales.

One observation that we did not anticipate is the hard downward trend toward the end of the sales dataset across all items. This reflects poorly on the future of these stores and requires further investigation.

Future Work

- Refine models and seek out other methods to acquire a more accurate prediction.
- Apply evaluation metrics to discover features of our models such as precision and specificity.
- Investigate items of other types that we suspect to have stable sales, such as sales data for CD recordings of well-established bands like AC/DC.
- Automate the model to predict across all items for each store.

References

1C Company. “Predict Future Sales.” *Kaggle*, 18 February 2018, https://www.kaggle.com/c/competitive-data-science-predict-future-sales/overview. Accessed 12 February 2021.

Coghlan, Avril. “Using R for Time Series Analysis.” *A Little Book of R for Time Series*, 2010, a-little-book-of-r-for-time-series.readthedocs.io/en/latest/src/timeseries.html. Accessed 29 March 2021.

Singh, Gurchetan. “Time Series Forecasting: Various Forecasting Techniques.” *Analytics Vidhya*, 23 Dec. 2020, www.analyticsvidhya.com/blog/2018/02/time-series-forecasting-methods/. Accessed 27 March 2021.