Proposal for Kids on 45th

Gary Gregg
Jahnavi Jasti
Abhishek Varma

Kids Helping Kids on 45th
Department of Data Sciences
University of Washington, 2017/2018

Submitted to:
Bookis and Elise Worthy
Kids on 45th

Table of Contents

Executive Summary

In the sections that follow, this proposal may refer to *Kids on 45th* as "the client," "the enterprise," or "the sponsor." We may refer to our own team, *Kids Helping Kids on 45th* as "we," or "the team."

Kids on 45th is a mixed retail and consignment shop in the Wallingford neighborhood of Seattle, Washington. The store also has an on-line presence. The business sells clothes and toys for children. The toys are typically new, but the clothing items are largely used. The majority of clothing is sold on consignment. Individual clothing items are sold in the retail store, or on-line, but not both. Usually the best best conditioned, name-branded and "photogenic" clothing items are sold on-line. The remainder is sold in the store. Less than 10% of the clothing items qualify to be sold on-line. The typical *Kids on 45th* customer has growing children, and children outgrow clothing. As their children grow, parents will attempt to sell clothing items by consignment. *Kids on 45th* says this about their customers:

> *We think of our consignors as friends and partners, working together to sell items quickly and for the best price.*

Upon item sale, *Kids on 45th* consignor partners receive 40% of the proceeds, available as a store or on-line credit. In this way, the partners may outfit their children with an ongoing stream of quality used clothing. *Kids on 45th* currently has more willing consignor partners than they can handle. Nevertheless, one challenge for this enterprise is the availability of quality new clothing that can be purchased from retail chains such as Target.

Bookis and Elise Worthy acquired *Kids on 45th* in June 2017, and inherited along with the business a database of legacy sales data with its origin from the late 1980s. When items are consigned to the store for sale, their associates will attempt to price items using intuition, judgment, and knowledge gained by experience. If items do not sell within a reasonable amount of time, they may be consigned to a bargain-bin status ($1.00 sale price), and after 30 days the item may be donated to charity.

Statement of Problem

When Bookis and Elise Worthy acquired Kids on 45th, they started a new sales tracking system that includes the following tools:

- Apache Lucene – Apache Lucene provides Java-based indexing and search technology
- Elasticsearch – Elasticsearch is an open-source, RESTful distributed search and analytics engine build on Apache Lucene
- Kibana – Kibana is an open-source data visualization and exploration tool used for log and time series analytics, application monitoring, and operational intelligence use cases

The legacy sales data – encompassing data from the late 1980s through 2017 – is contained in Microsoft Access tables. It is currently not in a form convenient for easy analysis by the enterprise.

In order to maximize revenue and keep the business as healthy as possible, it is important for *Kids on 45th* to price incoming consignment items correctly. Factors affected by pricing include total revenue for the store, customer satisfaction, and consignor partner satisfaction. These three factors together ultimately affect the success of the enterprise.

In order to better understand whether they are pricing items correctly, *Kids on 45th* requires an analysis of the legacy sales data that was left by the previous owners. Elise and Bookis Worthy are most interested in the question: *What Sells Things?* To this end, their first interest is a report with a detailed analysis of the legacy sales data.

Design Objectives

The primary objective of our team, *Kids Helping Kids on 45th* will be to assist the enterprise by analyzing their legacy sales data, and produce a detailed report with our findings. In order to fully address the needs of the enterprise, *Kids Helping Kids on 45th* will construct one or more models that can predict sales price for consignment items. These models make take into account item categorization (sweater, pants, shirt, shoes, etc.), branding, condition, quality and size. As a stretch goal, it may be possible to import the legacy sales data into the new sales tracking system that *Kids on 45th* has adopted.

To summarize, the desired deliverables include, in priority order:

1. A report with a detailed analysis of the legacy sales data

2. Model(s) that predict sale price based on item type, brand, condition, quality and/or size

3. Importation of legacy sales into the new sales tracking technology

Since *Kids on 45th* is an ongoing concern in the Wallingford neighborhood of Seattle, it may be possible for teams from future cohorts of the *Master of Science in Data Science* program at the University of Washington to build on the work of *Kids Helping Kids on 45th* to further assist this enterprise.

Technical Approach

Identifying Needs of Customer

Our team recognizes that regular communication with the ownership and management of *Kids on 45th* is an essential requirement for this project. We need this both to understand the ongoing needs of the enterprise, and to make certain that we remain on the right track as we proceed with our analysis of the legacy sales data. At the close of 2017, *Kids Helping Kids on 45th* has had three face-to-face meetings with Bookis and/or Elise Worthy, and more regular meetings are planed. The teams believes a weekly face-to-face meeting may produce optimal results. Most conveniently, these meetings can be held informally over coffee, and in close proximity to the retail store. Additionally, electronic or telephonic communication is easily available for needs that do not require a face-to-face meeting. At the close of 2017, both our team, and *Kids on 45th* ownership feel comfortable with each other, and we anticipate no problems going forward. See the **Communication with Sponsor** section in **Project Management**, below.

Identifying Target Specifications

Kids Helping Kids on 45th has done a preliminary analysis of the legacy sales data contained in Microsoft Access tables. In our *Data Pipeline Document*, we note that the legacy sales data of *Kids on 45th* is stored in four Microsoft Access databases (*.mdb extension). We reproduce the pipeline data here. The databases are named *Custdata.mdb*, *Product.mdb*, *Sales.mdb* and *Scandata.mdb*.

The **Custdata** database has seven tables: ‘Users’, ‘Sales’, ‘Maillist Profiles’, ‘Hold’, ‘Customers_Finance’, ‘Customers’, and ‘Cnwa’.

The **Product** database has six tables: ‘ProductsBu04Older’, ‘PP_Presets’, ‘PP_Descriptions’, ‘PP_Catagories’, ‘PP_Active_Products’, and ‘Active Back-Up 3600’.

The **Sales** database has seven tables: ‘Transaction Types’, ‘State Sales Tax’, ‘Sales’, ‘PP_Soldprod2008’, ‘PP_Sold_Products’, ‘PP_Returned_Products’, and ‘Gift_Cirt’.

The **Scandata** database has two tables: ‘CC_Descriptions’, and ‘CC_Con Check In’.

Note: The structure, meaning, and organization of the data is still under analysis. We document here only sources and data point counts.

Analysis of Data Points

Record counts for tables in the **Custdata** database:

Table Name	Record Count
Users	10
Sales	209,878
Maillist Profiles	1
Hold	1
Customers_Finance	8,793
Customers	8,816
Cnwa	1,904

Record counts for tables in the **Product** database:

Table Name	Record Count
ProductsBu04Older	29,223
PP_Presets	279
PP_Descriptions	4
PP_Categories	5
PP_Active_Products	147,563
Active Back-Up 3600	17,941

Record counts for tables in the **Sales** database:

Table Name	Record Count
Transaction Types	6
State Sales Tax	3
Sales	210,819
PP_Soldprod2008	76,587
PP_Sold_Products	538,024
PP_Returned_Products	0
Gift_Cirt	0

Record counts for tables in the **Scandata** database:

Table Name	Record Count
CC_Descriptions	19
CC_Con Check In	106,303

Analysis of Data Location

The legacy sales data has been shared with *Kids Helping Kids on 45th* team members by store owner Bookis Worthy using **Dropbox**. The legacy sales data is not currently being updated, and we can safely assume the data is in its final form. The data has been downloaded to team member laptops in order to perform the analysis required for the pipeline document, and this proposal document.

Software for Preliminary Analysis

Kids Helping Kids on 45th plans on using using a variety of platforms to analyze the legacy sales data (see below). For the preliminary analysis of data points, we relied on the following tools to access the data. For Linux systems, we used GNOME MDB. For Apple systems, we relied on Microsoft Access for Mac, and for Windows systems we relied on Microsoft Access for Windows.

Literature Review

We give in the **References** section of this document links for the on-line presence of *Kids on 45th*, links to the technology currently being used since June 2017 for sales data at *Kids on 45th*, links for the Microsoft Access database technology used to store the legacy sales data, and links to the technologies *Kids Helping Kids on 45th* shall use to access and analyze the legacy sales data.

Design Concept

For a Data Analysis Report: After continued consultation with our client, *Kids Helping Kids on 45th* believes that a deliverable consisting of a report of exploratory data analysis will provide immediate and actionable insights. We plan on cleansing the data by consolidating categories of items sold, dropping outliers and displaying the average (and/or median) prices of items. We propose to determine standard deviation provided for premium and discounted items. This will be amortized over time spent on shelf so that the client can account for this statistic with respect to price. It will allow our team to measure price elasticity of these goods. According to the client, unbranded goods are essentially commodities, and the price elasticity of demand for commoditized goods is very inelastic. This means that the demand for such goods is very sensitive to price. With such small margins, the ability to give a clear picture of optimal price will allow revenue optimization.

For a Pricing Model: *Kids Helping Kids on 45th* believes it may require assistance from Data Science faculty at the University of Washington to better understand how time-to-sale is influenced, and may be modeled by adjusting price. The team additionally believes that the optimal sales price model will be a straightforward, predictive linear model, or *Ridge Regression* model available in the *sklearn* Python package. These tools are also available in the R programming language. Additionally, we may use LASSO techniques to better understand what predictors drive sales price.

Kids Helping Kids on 45th proposes to use the following tools to analyze the legacy sales data, and develop a model to assist in pricing of items received for sale on consignment. See links for each of the below-named technologies in the **References** section of this document:

- R. While we anticipate all final models to be Python-based, we may use R for cleansing and analysis of data, as well as constructing sample models to determine model quality

- R Studio. R Studio provides a high-quality integrated development environment (IDE) for management of R language projects. It also has tools for debugging, and visualization of programming output

- Python. Python has become a programming language standard for construction and analysis of data

- Bokeh. Boken is a Python-based visualization library that targets modern web browsers for presentation purposes

- Pandas. Pandas is a Python library that provides high-performance, easy-to-use data structures and data analysis tools for the Python programming language

- Sklearn. Sklearn is a machine learning library for the Python programming language, featuring various classification, regression and clustering algorithms including support vector machines

- Pycharm. Pycharm is an easy-to-use IDE for Python development. It assists in code formatting, static analysis, and debugging

- Jupyter. Jupyter Notebooks are a convenient and powerful means to bring together programming cells in Python or R, and combine them with markup language cells for compelling and understandable presentations

Project Management

Deliverables

Kids Helping Kids on 45th proposes to provide to the sponsor the following deliverables by mid-March, 2018:

1. A report detailing a data-driven analysis of the relationship of price to expected rack-time (or time that an item is held in inventory), and how pricing policy can be instituted to maximize profitability.

2. A report (perhaps a combined report with that named in #1) that uses data-drive analysis to draw conclusions about item pricing, and pricing policies that drive sales. See the **Statement of Problem** section for the key question: *What Sells Things?*

3. A Python drive model that utilizes descriptive predictors for a consignment item to suggest a retail price for that item.

As time allows, we propose to import legacy sales data to *Kids on 45th* new sales analysis tools, *Elasticsearch* and *Kibana*.

Communication with Sponsor

Kids Helping Kids on 45th proposes weekly (or less frequently, as our sponsor desires) meetings over coffee at *Chocolati*, which is an establishment next to the Kids on 45th retail store. If another meeting location is desirable for an individual meeting or group of meetings, the team can accommodate the request as required. See the **Identifying Needs of Customer** section in **Technical Approach**, above.

Team Organization

Kids Helping Kids on 45th anticipates having a fluid leadership structure, with decisions made by consensus, or majority-rule. We propose to consult regularly with our project sponsor, *Kids on 45th*, and program coordinator at the University of Washington. We will divide tasks into three (or more) subsections, and attempt as much as possible to assign subsections to the individual with the best strengths in that area. Programming, data visualization, data cleaning, documentation are all relevant subsections. Regular contact with the sponsor for questions and clarification will be a regular feature to ensure that we remain on track, and on time.

The retail store of *Kids on 45th* is located in proximity to the residences of two of the team members. Regular face-to-face contact is feasible, and convenient. Telephone and electronic communication can supplement this manner of regular communication. See the See the **Communication with Sponsor** section, below.

Inter-team discussion and tie-breaking will be considered the optimal technique for conflict resolution, however, we believe resolution by the sponsor or University of Washington program coordinator may be required for more intractable disagreements. We hope this will prove unnecessary, and will work to avoid it.

Team Qualifications

See Appendix A, below, for complete résumés for each team member.

- **Gary Gregg –** skills: 20 plus year's experience as software developer, skilled at coding in multiple languages, native English speaker, skilled at writing. Gary has final say in software engineering disputes.

- **Abhishek Varma –** skills: Experience as a management consultant and project manager, requirements gathering and stakeholder analysis. Abhishek oversees requirements planning, and has final say.

- **Jahnavi Jasti –** skills: good at programming, data manipulating and creating visualizations in R, PowerBI. Jahnavi oversees visualizations, and has final say.

Conclusion

A preliminary analysis by the team of the legacy sales data of *Kids on 45th* indicates that classification of sold items by type, branding, condition, quality and/or size may be complicated by incomplete, or insufficient data. However, we believe that with we can work with the data that is present, and produce a high-quality analysis, and report that will assist the enterprise in optimal pricing of their consignment items. Adjustments for legacy sales data may need to take inflation into account as a factor for pricing in 2018, and years that follow. Inflation is a tricky subject. It is not determined at this time whether we shall use simple Consumer Price Indexing (CPI), or also increases in store rent and employee compensation. These last two factors may not be fully reflected by increases in baseline CPI, particularly for the Seattle area.

As described above, we anticipate models that the team will produce shall be straightforward linear, or *Ridge Regression* models.

References

- Access: https://products.office.com/en-US/access?legRedir=true&CorrelationId=37be4ac7-c838-4911-b97f-e693870a729e
- Bokeh: https://bokeh.pydata.org/en/latest/
- Dropbox: https://www.dropbox.com/?landing=dbv2
- Elasticsearch: https://www.elastic.co/products/elasticsearch
- GNOME DB: http://www.gnome-db.org/
- Jupyter: http://jupyter.org/
- Kibana: https://www.elastic.co/products/kibana
- Kids on 45th: https://kidson45th.com
- Lucene: https://lucene.apache.org/
- Pandas: https://pandas.pydata.org/
- Pycharm: https://www.jetbrains.com/pycharm/
- Python: https://www.python.org/
- R: https://www.r-project.org/
- R Studio: https://www.rstudio.com/
- Sklearn: http://scikit-learn.org/stable/

Appendix 1: Résumés of Team Members

Gary Gregg

"Inventive software engineer, with broad development skills across global environments, designs and develops efficient algorithms to solve complex problems and creates intuitive user interfaces for Linux, Android and iOS operating systems. Talented developer, with experience including mobile app and software development and defense contracting, is frequently recognized for high-volume, bug-free code production in high-visibility, time-sensitive situations." - Summary statement from CV.

Academic background:

- Bachelor of Science in Mathematical Sciences, emphasis in Operations Research, **University of California, Santa Barbara**
- Master of Science in Computer Science, **University of California, Santa Barbara**
- Second year graduate student, **University of Washington** with objective of Master of Science degree in Data Science, March, 2018

Professional background:

- Software Design Engineer, **Microsoft Corporation**, 1997 – 2000. Assigned to develop test software for the Windows operating system
- Embedded Software Engineer, **The Boeing Company** Defense & Space Division, 2002 – 2010. Assigned to X-45 UCAV, Unmanned Systems, P-8 Mission Systems Software, and Airborne Early Warning & Control (AWA&C) for the Republic of Korea
- Software Engineer, **The Boeing Company** Commercial Aircraft, 2010 – 2013. Assigned to Loadable Software
- Android Software Engineer, **Airbiquity, Inc.**, 2015 – 2016. Contract position to develop mobile software for connected car technologies

Skills: C, C++, Java, C#, Objective-C, Python, SQL, Visual Studio, Eclipse, XCode, Linux, Windows, Android, OSX, iOS, GIT, Subversion, Microsoft Access and SQL Server, Oracle, Sybase, Jira and much more

Differentiator: *"Operations Research is a discipline that deals with the application of advanced analytical methods to help make better decisions..."* (Wikipedia). Additionally, much of my graduate work in computer science focused on efficient algorithms to solve complex problem. **Optimal decision making, algorithmically and otherwise, is my passion.** I bring this same passion and expertise to the field of Data Science. I am a skilled programmer, bringing academic work and more than 20 years of experience to this new discipline.

Interests: **My particular field of interest is the sciences, to include both physical sciences and life sciences.** The effects of human activity on Earth's biosphere is of great interest to me. I would love to use my skills as a data scientist to prove the causes of climate change, and find ways to ameliorate its effects.

Abhishek Varma
Abhiv@uw.edu
4320 Aurora Ave, #N-201 Seattle, WA
713-628-4270

Education

University of Washington Master of Science
2016-2018
Data Science GPA 3.0

The University of Texas at Austin Bachelor of Business Administration **2014**
Management of Information Systems
Minor: Radio-Television-Film GPA: 3.3

Experience

Ernst & Young – Staff Consultant (Analytics); Los Angeles, California
2014 – 2016

- Conducted System Integration Testing for EY Helix, a $20m analytics initiative that employs Hadoop and Spotfire to visualize and generate accounting reports for large audit clients
- Conducted a Change Impact Analysis over the end to end supply chain of a major technology firm, assessing current and future states as well as determining the effects on stakeholders, policies, and processes
- Developed tools in MS Access that allowed a major pharmaceutical company to generate reports that matched existing controls and procedures with FDA regulations, helping identify gaps and uncover risks

Microsoft – Program Manager Intern; Redmond, Washington
Summer 2013

- Gathered requirements and designed UI for Capacityfolio dashboard after extensive interviews of stakeholders, empowering customers to make more informed and intelligent business decisions regarding server capacity
- Identified key pain-points in server capacity related escalation processes by mitigating ambiguity and risk and proposed potential solutions to IT leadership

Projects

Data Analytics – Box Office Earnings **Spring 2014**

- Analyzed Box Office earnings of successful films to find relationships between the genre and the release date of the film, allowing studios to find the optimum release dates to maximize revenue.
- Developed models using Python and Machine Learning libraries to conduct linear regression and SVM analysis

Entrepreneurship Experience

Internet Marketing – Multiple Sites **2015-2016**

- Developed multiple sites and drove traffic through various SEO and social media strategies
- Employed extensive analytics to understand user behavior and increase traffic and optimization
- Got 100k hits in a week of launch, with content being shared over 40k times on social media in 24 hours

Additional Information

Computer Skills: Python – SciKit and Pandas, R,SQL, Tableau, VB.Net, ASP.Net, ServiceNow (Admin), MS Access
Languages: English, Hindi (Native)
Interests: Filmmaking, Gaming, Soccer, Cricket