Recommendation Monkey

Bill.com

Quarter: Fall 2021

Professor: Jianwen Su

TA: Mason Corey

Team Name: Inspector Royale

Mentor: Stuart Ogawa

Company: Bill.com

Team Members:

Derek Garcia (Lead) - derekgarcia@ucsb.edu

Noa Kim (Scribe) - noakim@ucsb.edu

Carson Coley - carsoncoley@ucsb.edu

Jonas Zhang - huiyuzhang@ucsb.edu

Joe Zhuang - zhouzhuang@ucsb.edu

Introduction

Background:

If a business owner wanted to conduct real, statistical analysis on their transaction data they must either hire data scientists or do it themselves, both of which are sub-optimal options due to time and cost constraints. As a result, many small and large business owners eventually acquire useful financial data so large in volume that analysis becomes an overly difficult task.

There is currently no widespread method for analyzing smaller industries of a business archetype for investing purposes, but they do exist for larger businesses. Using Bayesian statistics and research data our project can offer a quantitatively and qualitatively supported recommendation to a business user.

Present Implementation:

SaleForce:

Business owners can manually enter their transaction data into a paid, cloud-based application so they can easily view analytical details and generate reports. However, this is largely impractical for any business user that must manually enter enormous amounts of financial data, and SaleForce often displays graphs but without any explainability.

HubSpot:

A free, dashboard style hub for tracking and managing customer data. While user-friendly and simple to use, HubSpot only serves as a way to manage customer data rather than analyzing and displaying it as comprehensible graphs.

Objectives, Implementation, Assumptions:

What is our objective? Who does it help and how?

> We aim to create a program that provides summarized reports on whatever industries that a business owner is interested in. The summarized report will display multiple, statistical graphs for each method used while effectively explaining what the user is seeing in business terms as opposed to confusing mathematics. The variety of statistical methods work together to provide a bigger picture of the industries' behavior, while also giving a definitive, actionable conclusion to the user so they can make informed business decisions. Business owners who have large amounts of financial data that they do not know exactly what to do to benefit most from this program.
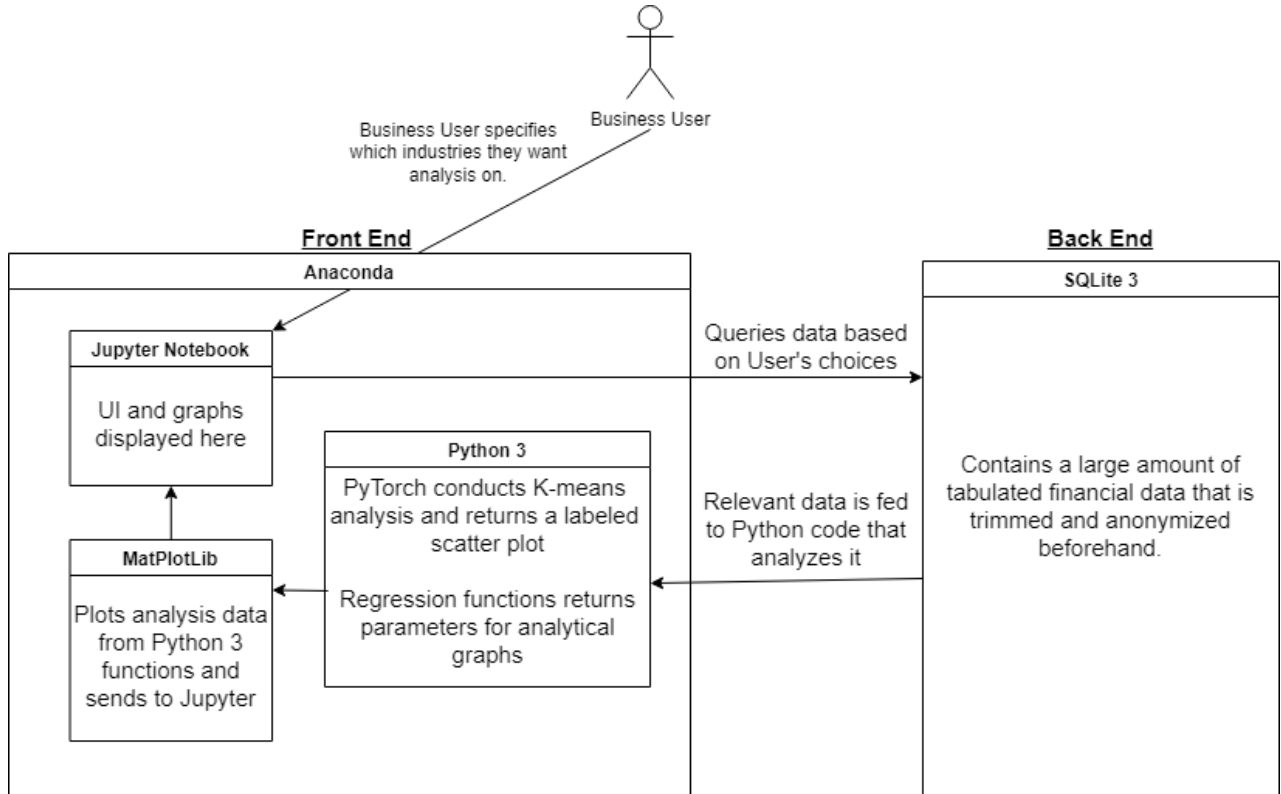
How will the solution be implemented?

> Jupyter Notebook will encompass our front-end and display a summarized report for the business user regarding the selection of industries they made, with graphs calculated by Python functions and plotted by MatPlotLib, along with text explanations of the observations and conclusions. Jupyter will first query the relevant data from the SQLite 3 database, which is fed to Python functions that will calculate the parameters for statistical graphs that MatPlotLib displays. Texts explaining the graphs and math conducted to the business user are added afterwards, based on what the dev team observed themselves about those industries.
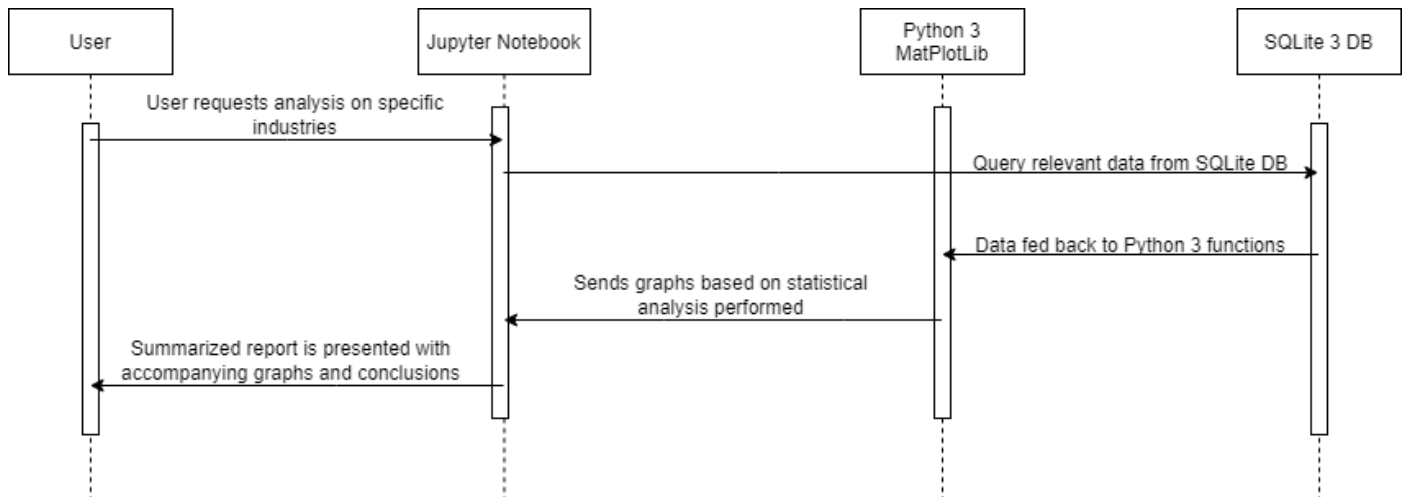
What assumptions had to be made?

> This program is strictly for business users that have access to large amounts of financial data in some database such that the Jupyter Notebook can successfully query the relevant data. The assumed format of this data is that the transaction date, sender id, receiver id, industry NAICS2 code, transaction amount, and company_size are given for each entity.
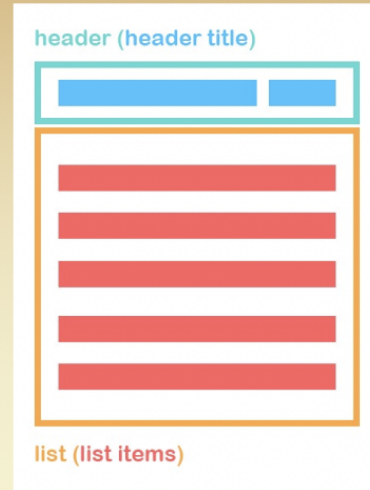
Diagrams

System Architecture Overview:

Business User

Business User specifies which industries they want analysis on.

Front End

Anaconda

Jupyter Notebook

UI and graphs displayed here

MatPlotLib

Plots analysis data from Python 3 functions and sends to Jupyter

Python 3

PyTorch conducts K-means analysis and returns a labeled scatter plot

Regression functions returns parameters for analytical graphs

Queries data based on User's choices

Relevant data is fed to Python code that analyzes it

Back End

SQLite 3

Contains a large amount of tabulated financial data that is trimmed and anonymized beforehand.

User Sequence Diagram:

User

Jupyter Notebook

Python 3 MatPlotLib

SQLite 3 DB

User requests analysis on specific industries

Query relevant data from SQLite DB

Data fed back to Python 3 functions

Sends graphs based on statistical analysis performed

Summarized report is presented with accompanying graphs and conclusions

UI Design:

Please select the

number of industries within the archetype **archetype_name**

Feature	Explanation
Yield	Amount of grain per unit of land
heck yield	Commercial variety's yield used as performanc
ld difference	Difference between yield and check y
Year	Year in which the variety was plante
anting date	Planting date
phic coordinates	GPS latitude and longitude of the farm or
emperature	Sum of daily temperatures in C in growing
recipitation	Sum of daily precipitations in mm in growin
lar radiation	Sum of daily solar radiation in W/m^2 in growi
CEC	Cation exchange capacity
pH	log of H+ concentration in soil
ganic matter	Percentage of soil made up of organic m
Soil class	Soil class category
Clay	Percentage of clay in soil
Silt	Percentage of silt in soil
Sand	Percentage of sand in soil
Area	Probability of growing soybeans in the su
PI	Soil productivity index

ilable in the Experiment dataset are listed in the left column, while their expla
e right.

10.1371/journal.pone.0184198.t001

Simple TAM Customer Report

Thesis and recommendation:

Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua. Ut enim ad minim veniam, quis nostrud exercitation ullamco laboris nisi ut aliquip ex ea commodo consequat.

Top Businesses to Pursue

1. Amazon
2. Google
3. Netflix

Date	Entry	Size	Close	Difference	% Return	P&L	Stdev
21-May-18	27.57	500					
24-May-18			27.8	0.23	0.83%	$115.00	
1-Jun-18			27.34	-0.23	-0.83%	-$115.00	
7-Jun-18			28.87	1.30	4.72%	$850.00	
16-Jun-18			30.1	2.53	9.18%	$1,265.00	
21-Jun-18			29.74	2.17	7.87%	$1,385.00	
28-Jun-18			27.8	0.23	0.83%	$115.00	
6-Jul-18			29.49	1.92	6.96%	$960.00	
12-Jul-18			29.33	1.76	6.38%	$880.00	
19-Jul-18			29.66	2.09	7.59%	$1,045.00	
26-Jul-18			30.69	3.12	11.32%	$1,560.00	
2-Aug-18			30.58	3.01	10.92%	$1,505.00	
9-Aug-18			29.03	1.46	5.30%	$730.00	
16-Aug-18			28.18	0.61	2.21%	$305.00	
23-Aug-18			27.98	0.41	1.49%	$205.00	0.94

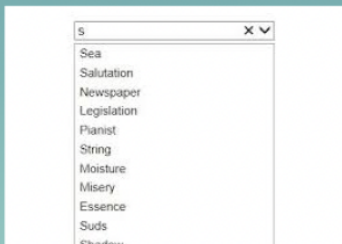Investopedia

🚩 Flag this report   More Info >>

Please select a simple or advanced customer:

simple | advanced

Simple Customer Report: Detailed Analysis

Top 20 Businesses

S
Sea
Salutation
Newspaper
Legislation
Pianist
String
Moisture
Misery
Essence
Suds
Shadow

Produce sales
IN THOUSANDS (USD)
OCT 2019 NOV DEC JAN 2020 FEB MAR APR MAY JUN

TAM:
Revenue Growth:
Activity:

Back to summary <<

Qualitative Research:
Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua.

Scores of Sectors:
Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua.

10 Key Attributes:
Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua.

User Stories / Use Cases

1. **As a business user, I want to view a cluster graph of transaction amount vs. time in a given industry so I can easily visualize similar transaction groups.**

    Acceptance Criteria: Graph appears plotting all transactions from the given industry with transaction amount as y-axis and transaction date as x-axis, with cluster centers plotted and corresponding color labels applied. Data is classified into three distinctly colored clusters using a K-means algorithm displaying layers of transactions.

    Scenario: User selects transaction vs time and industry and is shown a visual cluster graph of the transaction groups

    Github Issue: https://github.com/DerpDerek/InspectorRoyale/issues/12

2. **As a business user, I want to view a scatter plot of the transactions of the middle cluster from the aforementioned graph so I can better visualize transactions that are between non-significant and significant.**

    Acceptance Criteria: Graph appears plotting all transactions given by the middle cluster of the previous graph, with each transaction being colored according to the unique sender ID involved in that transaction. Linear regressions of each unique sender ID's translation can be viewed as well as the rates of change.

    Scenario: User selects scatter plot and is shown which transactions that are non-significant or significant.

    Github Issue: https://github.com/DerpDerek/InspectorRoyale/issues/13

3. **As a business user, I want to view a bar graph of the net transaction acceleration of each sender ID from the aforementioned middle cluster so I can visualize which senders enjoyed the highest increase to the rate of change of their transaction amounts.**

   Acceptance Criteria: A bar graph appears where each bar is colored and labeled along with a legend of the previous linear regression graph, and the height of each bar shows the net acceleration of the transaction amount.

   Scenario: User selects bar graph option and net transaction acceleration to see rates of change for transaction amounts

   Github Issue: https://github.com/DerpDerek/InspectorRoyale/issues/14

4. **As a business user, I would like to view the increment trend of total transaction amount or transaction count per day/month to know how positive each industry/company develops throughout the time**

   Acceptance Criteria: Graph of the approximation polynomial function should appear together with coefficients for each variable, and the function should expand to the corresponding date and the value on that date to be displayed.

   Scenario 1: User selects the order of derivatives of any natural number for the fit function

   Scenario 2: User selects a date in the future for the value prediction

   Github Issue: https://github.com/DerpDerek/InspectorRoyale/issues/15

5. **As a business user, I would like to view the periodic trend of total transaction amount/transaction count per day/month to know how the transactions of each industry/company fluctuate through the time.**

> **Acceptance Criteria:** The graph of the approximation sine function should appear together with the approximated coefficients for each variable, and the function will expand to the corresponding date and the value on that date will be displayed.
>
> **Scenario 1:** user inputs guessed parameters for the trigonometry function
>
> **Scenario 2:** user selects the a date in the future for the value prediction
>
> **Github Issue:** https://github.com/DerpDerek/InspectorRoyale/issues/16

6. **As a business user, I would like to generate an advanced summary TAM report so I can easily see details of transaction amounts about selected industries.**

> **Acceptance Criteria:** A summary report should show transaction amounts for specific industries. Unique sender and receivers should be shown and a graph with total transaction amount per month and average transaction amount for a selected industry.
>
> **Scenario 1:** User selects "Transaction Amount" option and filters for industries. A graph should appear showing total transaction amount per month and average transaction amount per month about selected industries
>
> **Scenario 2:** User searches for transaction amount with a specific industry. A graph should appear showing the number of unique senders and receivers in a given time period and their average transaction amount.

Github Issue: https://github.com/DerpDerek/InspectorRoyale/issues/17

7. **As a business user, I would like to view transaction details per day/month/quarter**

    Acceptance Criteria: Different filters can be chosen for a time period: day, month and quarter to get a line chart. A table of transactions can be filtered within a range, industry, date, and transaction amount.

    Scenario 1: User selects "Transaction Amount" option and filters for time period. A line chart should appear plotting max and min transaction amount per day/month/quarter

    Scenario 2: User selects "Transaction Amount" option and filters for certain average transaction dollar amount. A table should appear showing all senders and receivers that make transactions within the filtered range along with their industries, transaction dates, and average transaction amount.

    Github Issue: https://github.com/DerpDerek/InspectorRoyale/issues/18

8. **As a business user, I should see a sequential flow of numerous data visualizations**

    Acceptance Criteria: Data visualizations should include, 'All transaction history', 'Sum by day', 'Sum by Month', 'Regression Analysis', 'Number of transactions by day', etc, illustrating the future market for a given industry.

    Scenario 1: Users can select 'Regression Analysis' and see recommendations based on that method.

Scenario 2: Users can select 'All transaction history' to see and visualize the data given.

Github Issue: https://github.com/DerpDerek/InspectorRoyale/issues/19

9. **As a business user, I should be able to view each statistical approach used to calculate basic stats and realize, via static content, why the results are significant.**

Acceptance Criteria: Options to choose different statistical approaches like, basic calculations on the dataset, K-means clustering, and sine curve experiments/models

Scenario 1: Users can choose calculations on the dataset, to see and visualize the method on the data.

Scenario 2: Users can choose K-means clustering to see this specific method on the dataset.

Github Issue: https://github.com/DerpDerek/InspectorRoyale/issues/20

10. **As a business user, I should see different data and data analysis**

Acceptance Criteria: The different data analysis options that can be chosen include: Objective, Data Engineering Approach, Data Science Approach, K-means clustering, regression analysis and recommendation

Github Issue: https://github.com/DerpDerek/InspectorRoyale/issues/21

11. **As a business user, I want to view the top 20 businesses to pursue within an archetype**

Acceptance Criteria: Based on statistical methods, a list of businesses to pursue are seen based on an archetype and is sorted by the total preserved volume.

Scenario 1: User can select an archetype and is given a list of businesses

Github Issue: https://github.com/DerpDerek/InspectorRoyale/issues/22

aa

Technologies

- Anaconda:
    - Jupyter Notebook:
        - Front-end flow for customer presentation, in-line graphing and visualization
    - Python 3:
        - Numpy: support multi-dimensional arrays for data processing and analysis
        - Scipy: support methods like function-fit and other Math & Stat algorithms
        - Matplotlib: visualize data sets and processes of analysis
        - Pytorch: provide libraries for k-means and other machine learning algorithms
- SQLite
    - DB Browser:
        - Provide environment for efficient DB transactions and table manipulations