Food Recommender

By Ghodrat, Tejal, Shashank and Thomas

Why we need it ?

Because food is the most essential part of life.

In today's world we have endless number of choices for food.

To get desired food item customer has to go through a deep research e.g.read the ingredients, reviews, methods of preparation.

These endless number of choices confuses the c

"Food Recommender System" is made to make life people while choosing food products.

Tools

Git – version control

Node – Programming Environment

PostgreSQL – Database

Sqitch – Data Change Management

GIT

To push and pull code changes and keep the version history of the code.

NODE

Simple and fast. Need to know javascript to code

Libraries like express helps implement which HTTP requests we wish to respond and how we respond.

Postgresql

Open Source

Better Data Integrity

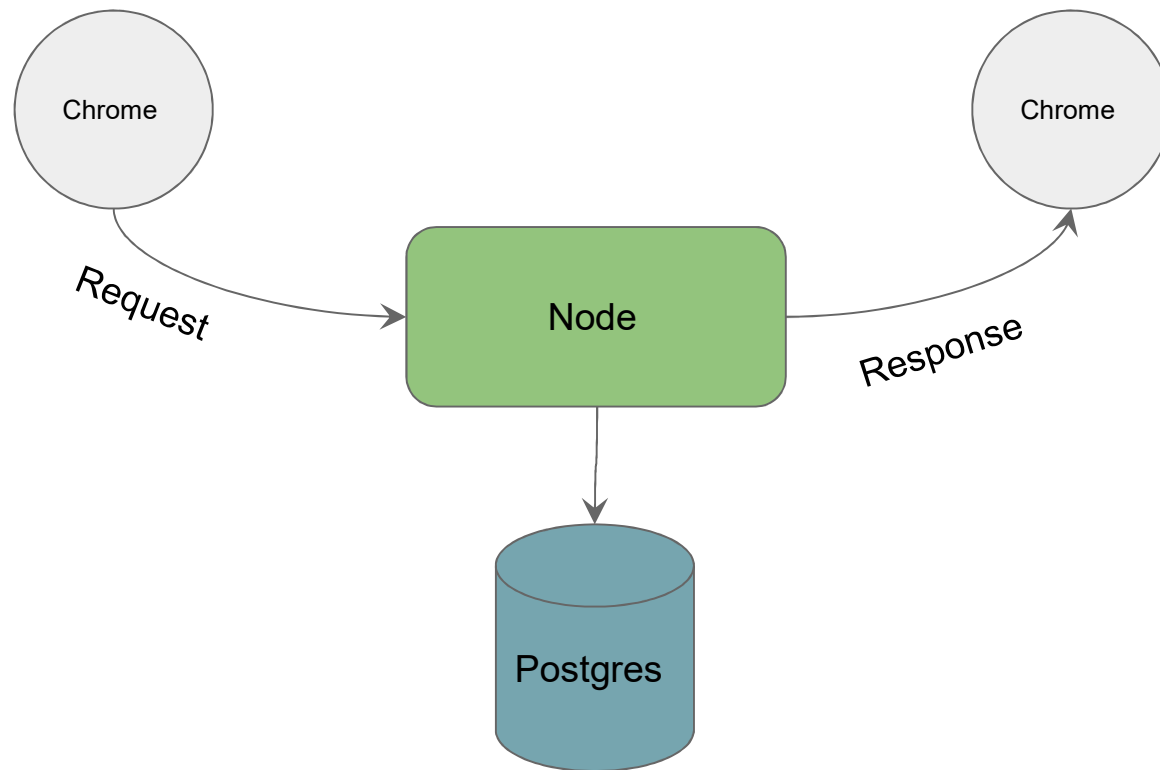Easily handles complex design

Sqitch

To manage changes in Schema we used Sqitch.

Sqitch uses three files: deploy, revert, verify.

It helped us a lot to improve our schema.

Web Application Architecture

Chrome

Chrome

Node

Request

Response

Postgres

Raw Data

We used Web Fine Food Dataset. This food set contains food
   items sold by Amazon.

Summary of the dataset :

Food Products : 74,258

Customers : 256,059

Reviews: 568,454

product/productId: B001E4KFG0
review/userId: A3SGXH7AUHU8GW
review/profileName: delmartian
review/helpfulness: 1/1
review/score: 5.0
review/time: 1303862400
review/summary: Good Quality Dog Food
review/text: I have bought several of the Vitality canned dog food products and have found them all to be of good quality. The product looks more like a stew than a processed meat and it smells better. My Labrador is finicky and she appreciates this product better than most.

Amazon Data Fetching and Scraping

Colorado Spice Company, Beef, Poultry, Pork and Lamb Spice, Carne Adobado (Red Chili) 1.5-Ounce Packet (Pack of 12) by Colorado Spice
Be the first to review this item

About the Product

- Made with garlic, tomatoes, apple cider vinegar, brown sugar, sea salt and spices
- Use to create Pork Red Chili or any delicious culinary concoction
- Perfect for delicious home-cooked meals made easy
- Contains no dyes, no artificial colors, and no MSG
- Made in Boulder, Colorado, USA

This item's packaging will indicate what is inside. To cover it, select **Ship in Amazon box** on the checkout page.

Price: **$23.95** ($1.33 / Ounce) ~~$25.21~~
Save $1.26 (5%)

In Stock. Ships from and sold by Amazon.com.

Subscribe & Save 5% 15%
$23.95 ($1.33 / Ounce) Add 5 items to this month's delivery to unlock 15% savings on this item. Learn more

One-time Purchase
$25.21 ($1.40 / Ounce) Prime

Deliver 1 every
1 month (Most common)

Subscribe now

Add to List

Share

Subscribe & Save up to 35% on Best-Sellers
Click here to shop now.

Product Description

Since 1988, Colorado Spice has been providing the finest restaurants, food manufacturers, and grocers throughout the U.S. with high-quality, all-natural and organic herbs and custom spice blends. Colorado Spice is one of the most respected spice companies in the nation! At Colorado Spice, we understand the importance of a home-cooked, delicious meal that is quick and easy to prepare. We are passionate about helping our customers prepare tasty, unique meals utilizing our flavorful spices. All of our spice blends include an easy-to-follow, easy-to-prepare, chef-developed recipe that is sure to wow! Our blends are authentic to the cuisine from which they are developed. They are unadulterated and pure as defined by the Pure Food and Drug Act. Our products have no dyes, no artificial colors, and no MSG. All of our products are produced in our Boulder, CO facility. The owner, Rodney Smith, spends countless hours sourcing the best possible herbs and spices throughout the U.S. and the world. The Colorado Spice single-use retail bags offer a great way to prepare exquisite gourmet meals at home. Each of our fifty blends include a protein specific recipe straight from our chef to your kitchen. We are proud to make it easy to produce delicious home cooked meals.

Product Details

Shipping Weight: 1.4 pounds (View shipping rates and policies)
Domestic Shipping: Currently, item can be shipped only within the U.S. and to APO/FPO addresses. For APO/FPO shipments, please check with the manufacturer regarding warranty and support issues.
International Shipping: This item is not eligible for international shipping. Learn More
ASIN: B000J3AX96
Item model number: 5363039
Average Customer Review: Be the first to review this item
Amazon Best Sellers Rank: #512,749 in Grocery & Gourmet Food (See Top 100 in Grocery & Gourmet Food)
#2709 in Grocery & Gourmet Food > Herbs, Spices & Seasonings > Mixed Spices & Seasonings > **Meat Seasoning**

Important Information

Ingredients
Garlic, Tomato Powder, Apple Cider Vinegar (Maltodextrin, Food Starch-modified and Apple Cider Vinegar), Brownulated Sugar (Brown Sugar, Cane Caramel Color), Sea Salt, Lime Juice Powder and Spices.

Directions
Carne Adobado (Pork Red Chili): Use 1.5 pounds of pork chops, blade steaks or boneless pork ribs and cut into 1/2 inch cubes. Simmer in pre-heated heavy skillet and brown. Pour off excess fat. Add spice mixture and 1/2 cup raisins, 3 large chopped tomatoes and simmer for 1 hour or until pork is very tender. Garnish with lime and serve with rice and tortillas.

Legal Disclaimer
Actual product packaging and materials may contain more and different information than what is shown on our website. We recommend that you do not rely solely on the information presented and that you always read labels, warnings, and directions before using or consuming a product. Please see our full disclaimer below.
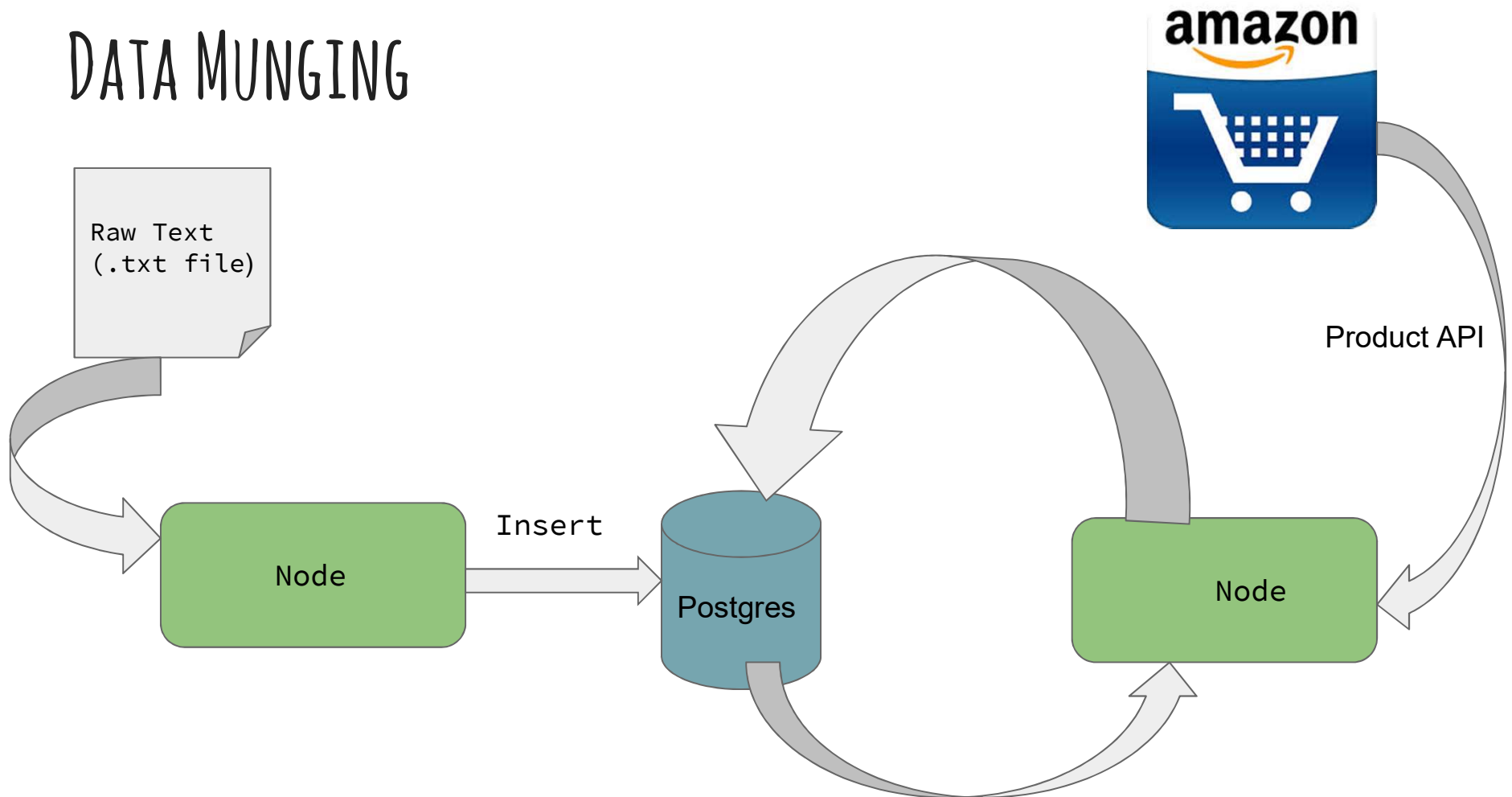
Normalized Data

product

asin(primary key)
B001E4KFG0

customer

id(primary key)	name
A3SGXH7AUHU8GW	delmartian

review

Product_asin (foreign key)	Customer_id (foreign key)	time	score	summary	body	helpfulness
B001E4KFG0	A3SGXH7AUHU8GW	1303862400	5.0	Good Qua...	I have bought several...	1/1

Data Munging

Raw Text
(.txt file)

Node

Insert

Postgres

Node

amazon

Product API

Data Analysis

For data analysis we used **Regression Models.**
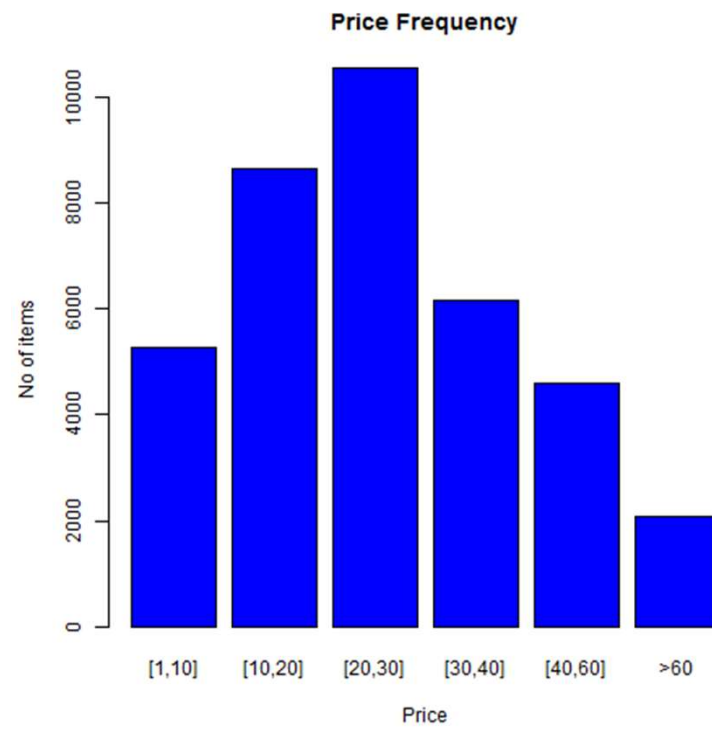Built in function of R is used to make
> ➢Histogram
> ➢Bar-Graph
> ➢Scatter-Plot

- These graph help us to visualize the data.
- **Collaborative Filtering** is used to find similar items and avoid unrelated items.

Data Analysis Architecture

HistoGram

Price Frequency

PIe-chart For Percentage of Reviews

Number of Reviews

🟥	4--10
🟨	10--20
🟩	20--30
🟦	30--40
🟦	40--50
🟪	>50

Products

56.1

7.2

2

3.5

8

23.2
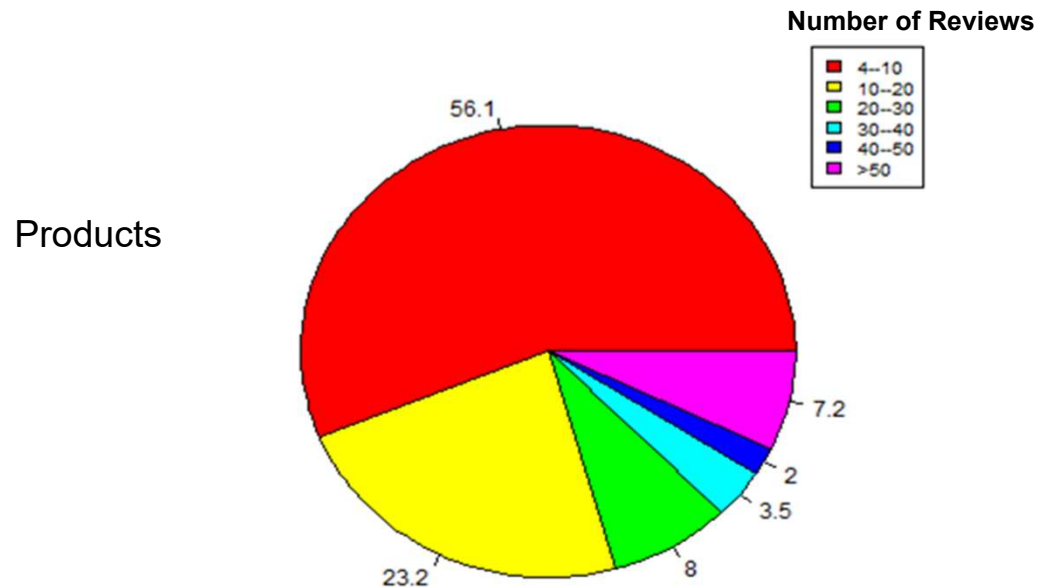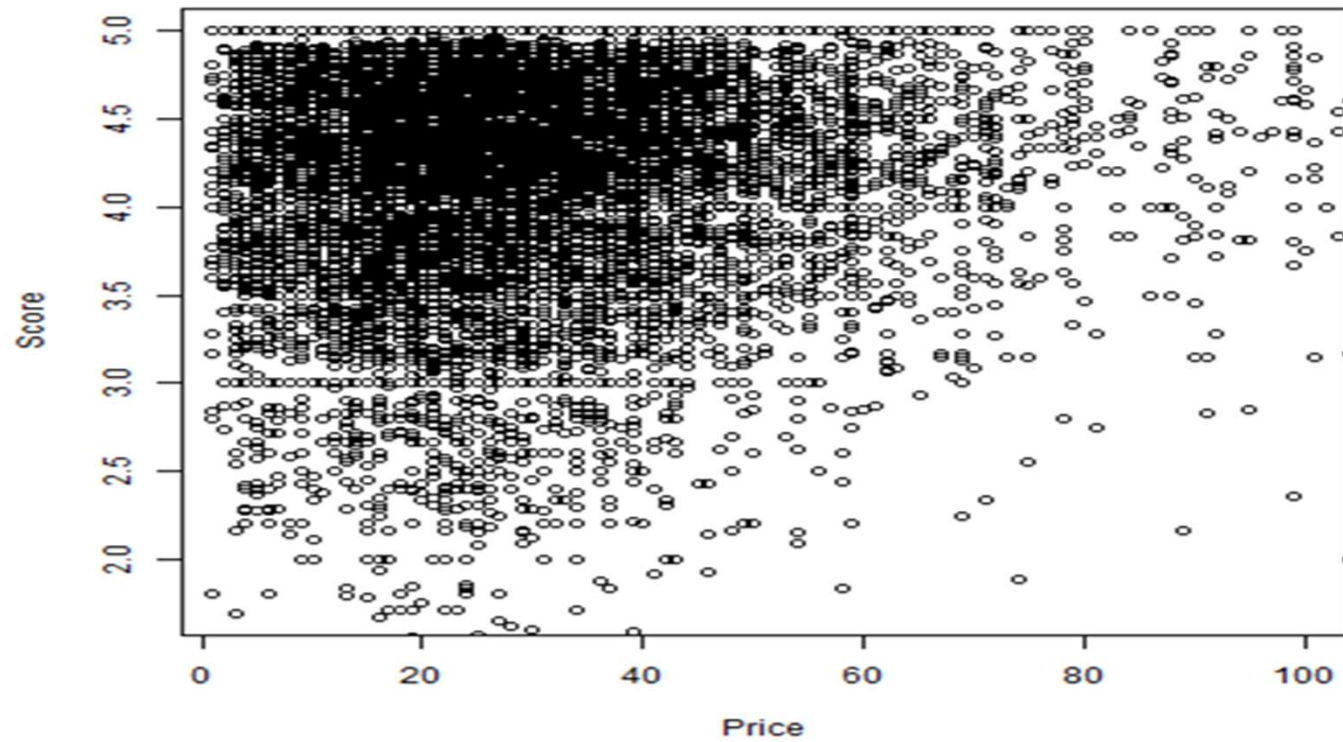
Half of our products have only 4-10 reviews

Reviews vs Price
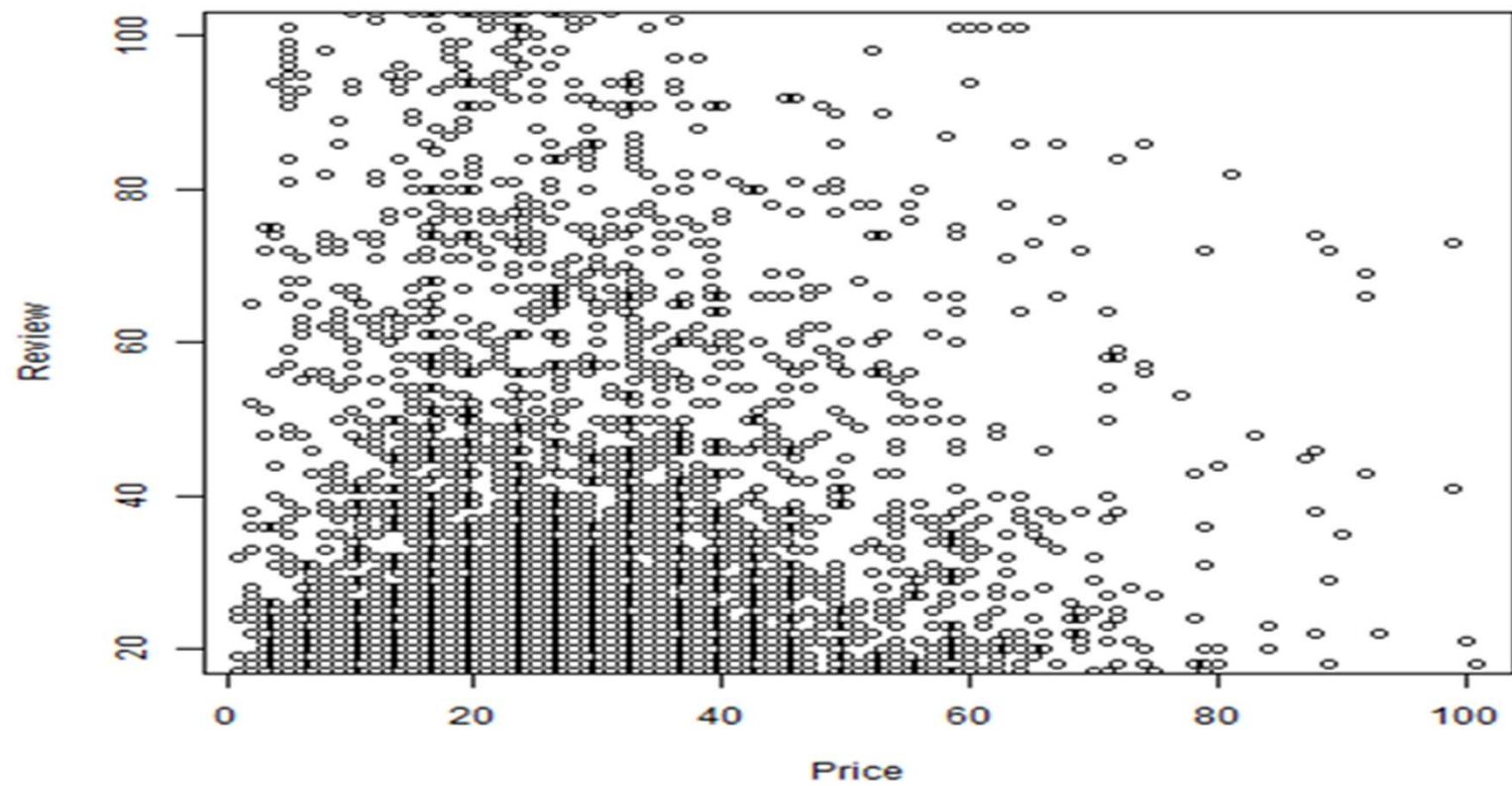
Score vs Price

Score Vs Number of Reviews
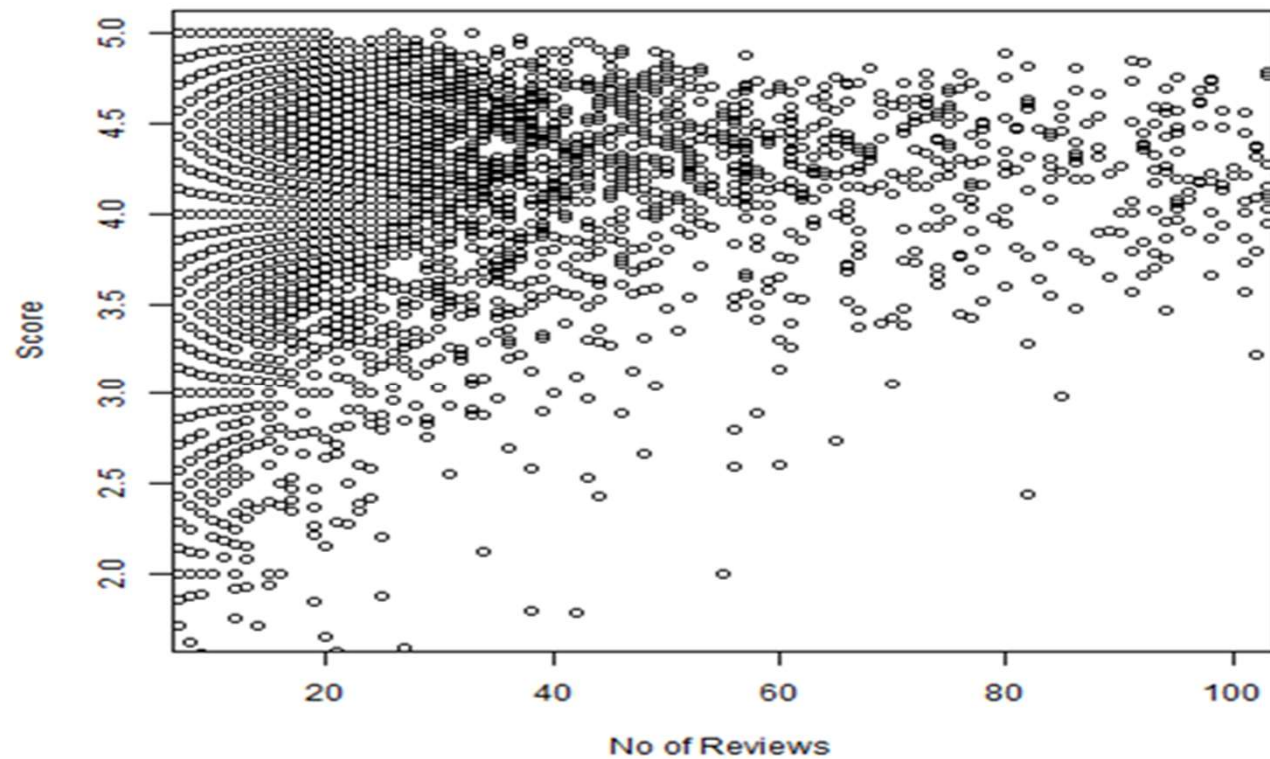
Food word Count

	A	B	C	D	E	F	G	H	I	J
1	and	199864								
2	of	183345								
3	the	143011								
4	a	106377								
5	to	95536								
6	in	88531								
7	with	75928								
8	for	67245								
9	is	65915								
10	are	40473								
11	Pack	40405								
12	or	38959								
13	your	30932								
14	from	28328								
15	that	27201								
16	you	25713								
17	The	25414								
18	flavor	23883								
19	our	23499								
20	as	22488								
21	Tea	19425								
22	natural	19405								
23	tea	19259								
24	on	17248								
25	by	17100								
26	Organic	16015								
27	taste	15322								
28	it	15305								

	A	B	C	D	E	F	G	H	I	J
	mono-floral									
112178	mono-pollen	1								
112179	monobasic-339	1								
112180	monoesters	1								
112181	monogrycerides	1								
112182	monohydrochloride.·	1								
112183	monoitrate(Vitamin	1								
112184	monoleate	1								
112185	monopotassico	1								
112186	monosaccharides	1								
112187	monosacchoride	1								
112188	monosacchride	1								
112189	monosaccride	1								
112190	monoster	1								
112191	monotonous	1								
112192	monotony	1								
112193	monsieur	1								
112194	monsoons	1								
112195	montebaldo	1								
112196	monterrey	1								
112197	month.</p	1								
112198	month/yyy	1								
112199	month</b	1								
112200	months-8	1								
112201	months-Milk-based	1								
112202	months-chocolate	1								
112203	months...</p	1								
112204	months. Phot	1								
112205	months. Textured	1								

Example

Tea	Sweet	Chocolate	Array_count_vector
0	1	5	[0,1,5]
0	0	0	[0,0,0]
1	3	4	[1,3,4]
.	.	.	....
		Average	[.......]

Mathematical Formula

Characteristic count vectors X_i.

If you have reviewed products $1, \ldots, n$, then consider

$$D_U = \frac{1}{n} \sum_{i=1}^{n} w_i X_i$$

$$w_i : \text{score}$$

Problem:

$$\text{Find } U : d(D_U, D) \text{ is min in } ||\ldots||_2$$

Recall : $d(X, Y) = ||X - Y||_2^2 = \sum_{i=1}^{n} (x_i - y_i)^2$

Implementation

```
CREATE OR REPLACE FUNCTION distance(real[], real[]) RETURNS real AS $$
DECLARE
  a1 ALIAS FOR $1;
  a2 ALIAS FOR $2;
  s real;
  BEGIN
    s := 0;
    FOR i IN array_lower(a1, 1)..array_upper(a1, 1) LOOP
      s := s + ((a1[i] - a2[i]) ^ 2);
    END LOOP;
  RETURN |/ s;
  END;
$$
LANGUAGE plpgsql;
```

Implementation

```
WITH dt as (
SELECT prc.*
  FROM big_prc prc
  WHERE prc.asin is not a seen item
  AND prc.score_avg > 3
  LIMIT 15000
)
SELECT asin, …, MIN(distance(user's count vector, customer_count_vector)) AS count_dist
FROM dt
GROUP BY asin, …
ORDER BY count_dist asc, score_avg desc
```

Successes

Programmatically generate giant SQL statement that can insert 10,000 reviews at a time, while concurrently scanning and parsing other data to be inserted.

Get to understand how R is useful in recommendation system.

Find out way to use Amazon HTTP API using Javascript.

Scrape data from amazon using http parser in JS.

Determine method to extract numeric values using word counts

Create numeric representation of customers using extracted word counts

Use collaborative filtering to match users to customers in the database.

Challenges

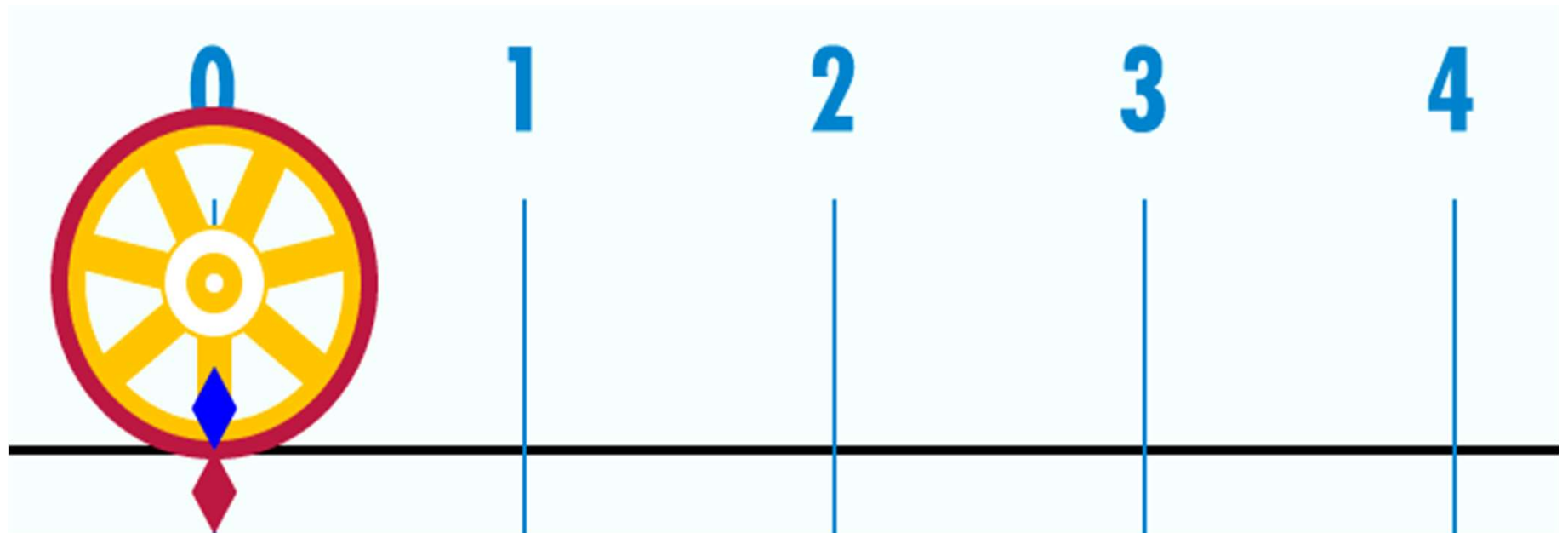Setting up tools for mac and windows

Using sqitch and switching from a local database to a centralized one.

Getting numeric values from data for data analysis

Learning New Tools.

Optimizing queries / denormalizing data for efficient searching and fetching
    from DB.

Demonstration

Thank You