

CS5200 DBMS  
Summer 2021  
The Debuggers  
PM5

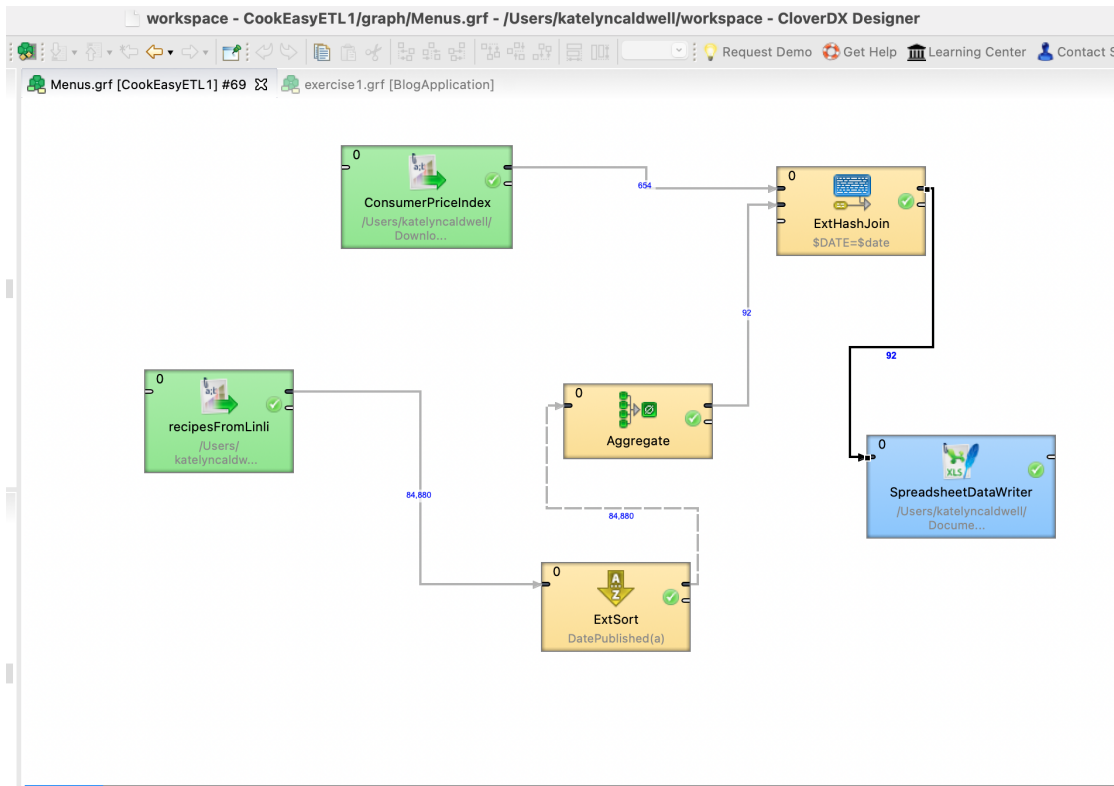
Yongjiang Yu  
Linli Mei  
Colin Nordquist  
Kate Caldwell  
Clara Jiaoyue Lan  
Saurav Shah

## Consumer Price Index of Food Across US Cities Over Time

### ETL #1

#### Data Source:

- Federal Reserve Board
- Data contains the US city average Consumer Price Index (CPI) for every month from January 1967 to present.
- We plan to use the data to examine whether there is any correlation between the CPI and number of recipes that were uploaded to Food.com over time. We are curious to examine whether more, less, or the same rate of recipes are being uploaded as CPI has increased.
- <https://fred.stlouisfed.org/series/CPIFABSL>



### ETL Workflow Components:

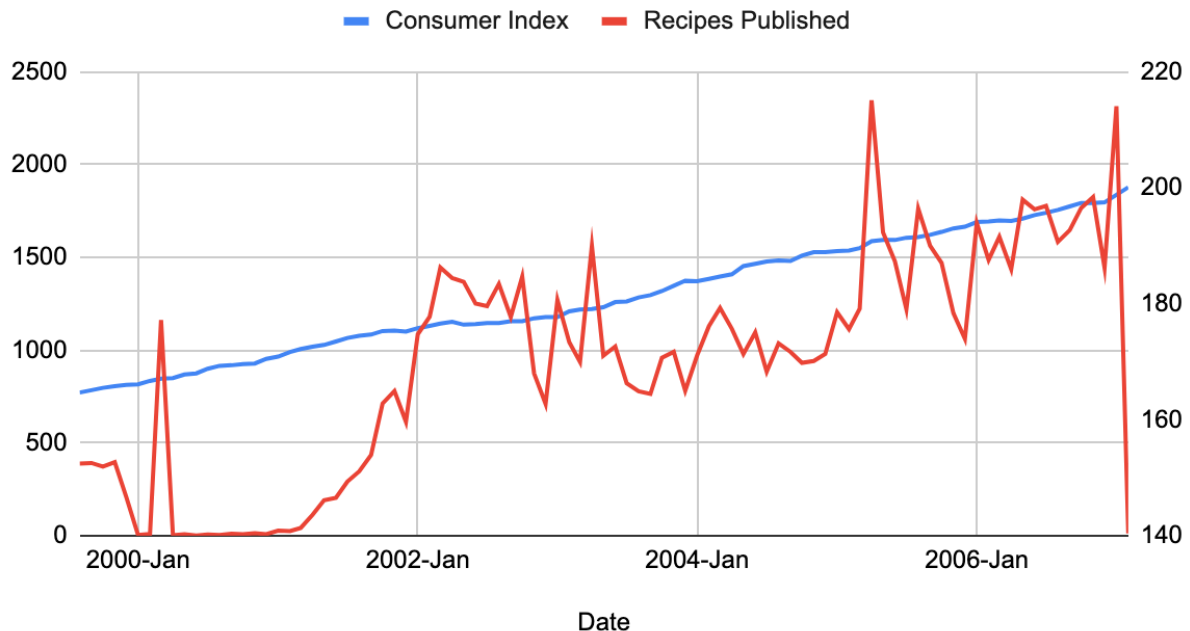
- **recipesFromLinli:** The CSV containing all the recipes used by our website, contains recipes and their relevant information such as Name, keywords, date created, ingredients, etc.
- **ExtSort:** used to sort the recipes by date
- **Aggregate:** Aggregates the recipes by date (specific to the Month), so that our data now contains the count of recipes created for each month/year
- **ConsumerPriceIndex:** the CPI data from the Federal Reserve Board containing data regarding the Consumer Price Index of food over time
- **ExtHashJoin:** INNER JOINS the two datasets ON the date (specific to the month)
- **SpreadsheetDataWriter:** writes the joined data to an excel spreadsheet

### Charts:

#### Chart #1

- **Hypothesis:** Our hypothesis for the outcome of CPI increasing over time on the number of recipes being uploaded over time to Food.com over time is that it will have no effect. The cost of food rising over time should have a strong correlation to the rate of inflation, and therefore cost the same to consumers in relative terms. However, another speculative possibility is that more recipes may be uploaded when CPI increases, as a cost-saving measure in response to prices of goods rising. As prices rise, restaurants will also increase prices and drive people towards cooking more at home.

## Relationship between Consumer Index and Recipes Published



### Interpretations from the Graph:

- Both CPI and recipes published show increases over time.
- Doesn't appear to be a particularly strong correlation, the number of recipes published in a year also appears much more volatile than the constant increase in CPI.

## Dishes Historically Included on Menus

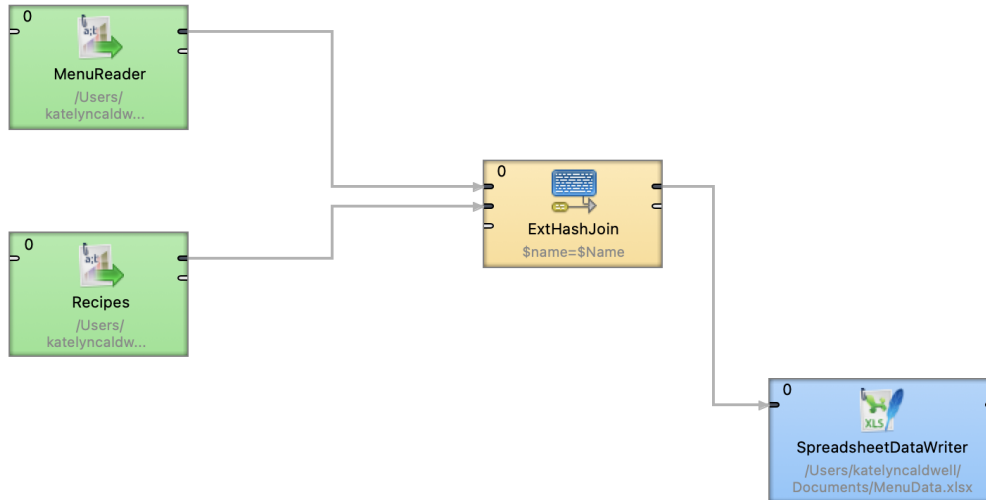
### ETL #2

#### Data Source:

- [data.world](http://data.world)
- Data contains historical restaurant menus dating back from the 1840s to 2010s. The data comes from the New York Library effort called "What's on the menu?"
- We plan to use this data to draw comparisons between different metrics to draw business insights that will help us create curated lists for our users and additionally help us determine what recipes are timeless and effectively would have the most attention on our website.
- We are planning to draw comparisons for the following 4 metrics:
  - Menu Appearances vs Review Counts
  - Menu Appearances vs Average Rating
  - Years Recipe has been on the Menu vs Review Counts

- Years Recipe has been on the Menu vs Average Rating

### ETL Workflow Components:

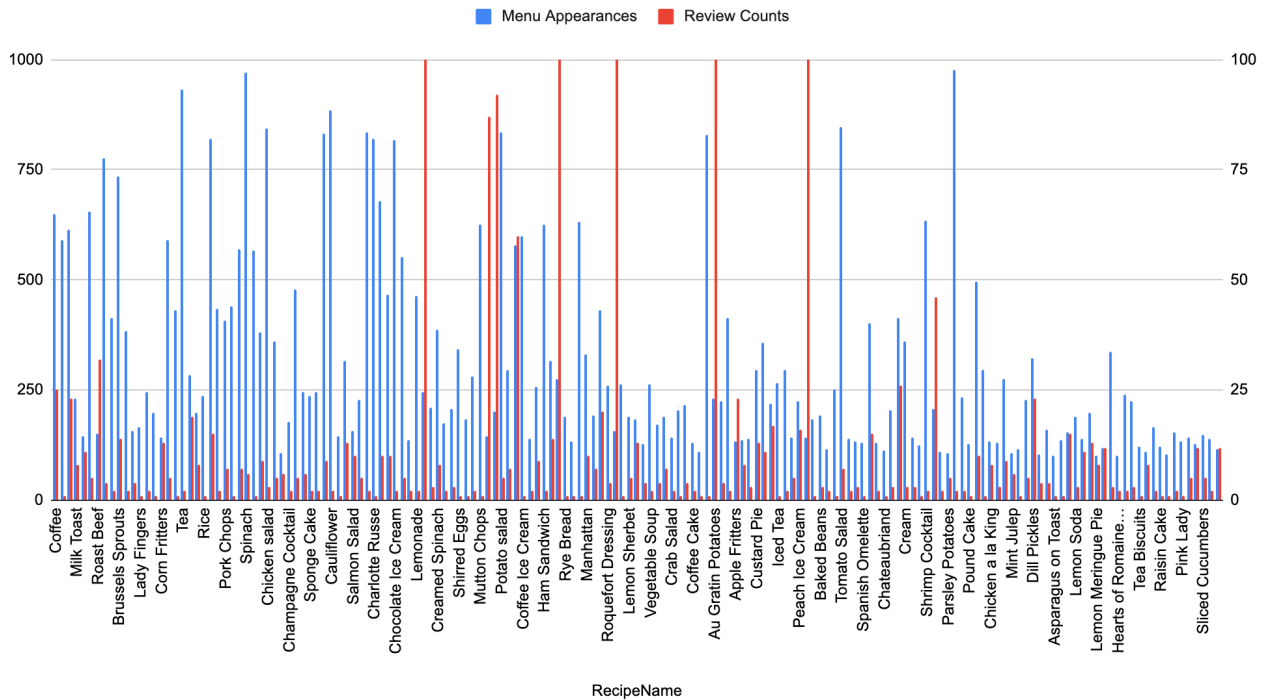


- MenuReader - Reads data from the Historical Menus dataset
- Recipes - Reads data from our Recipe Database
- ExtHashJoin - Joins the data on the Recipe Name
- SpreadsheetDataWriter - Transforms the data into a spreadsheet for our use to create charts.

### Chart #2:

- Hypothesis: This chart shows the trend between how many times a recipe has appeared on different menus vs how many times it has been reviewed on CookEasy. Our hypothesis would be that more appearances on menus would equate to more reviews as the recipes would be more classic in a sense.

## MenuAppearances and ReviewCounts



Quick Note: We have excluded recipes that appeared less than 100 times on menus to reduce the number of recipes and keep the most popular recipes amongst historical menus.

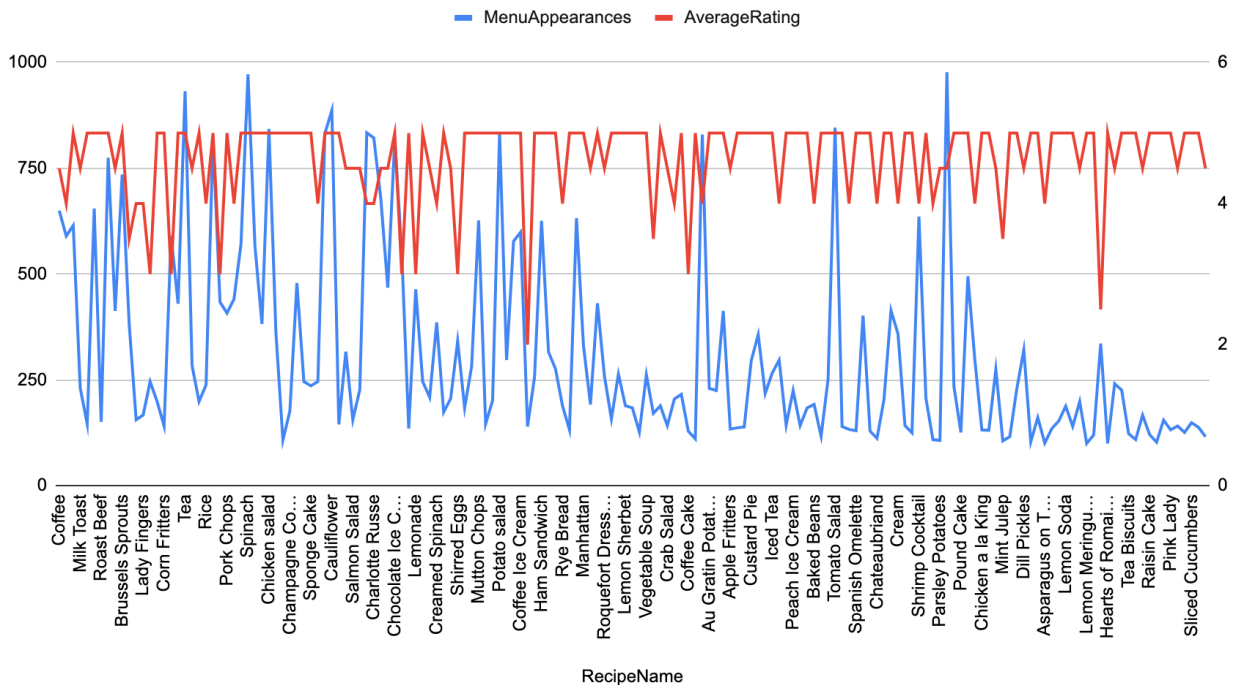
### Interpretation from the chart:

- There are a couple of conclusions we can make while interpreting this graph:
  - Some of the more popular recipes do not have enough reviews, we need to increase reviews on those as those will most probably be the most accessed recipes.
  - We can also see that some timeless recipes like Lemonade and very well reviewed and have a good amount of menu appearances as well, which shows that our hypothesis does hold for some recipes.

### Chart #3:

- Hypothesis: This chart shows the trend between how many times a recipe has appeared on different menus compared to the average rating by CookEasy users. Our hypothesis here is similar to the previous chart, more the number of appearances, the average rating would show us whether the popular recipe has a good user rating or not.

## MenuAppearances and AverageRating



Quick Note: We have excluded recipes that appeared less than 100 times on menus to reduce the number of recipes and keep the most popular recipes amongst historical menus.

### Interpretation from the Graph:

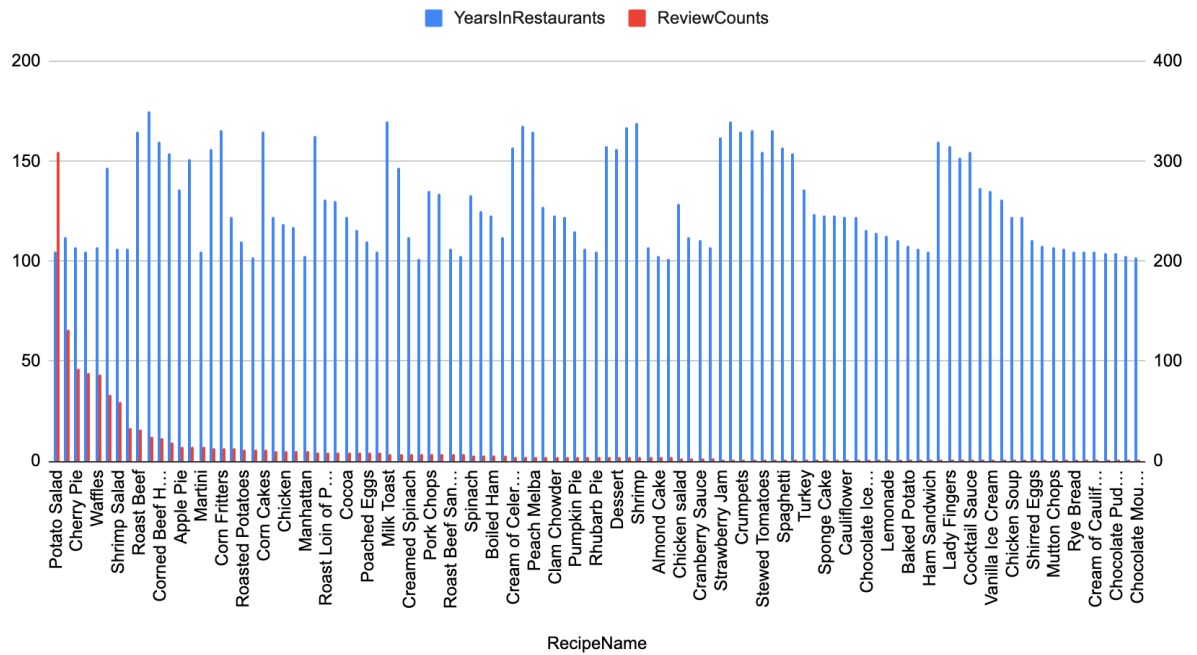
- We can conclude with this graph that the recipes that are popular according to the menu appearances indeed hold a good rating.
- The recipes above where the rating is less than 4, need to be looked at for quality purposes and maybe CookEasy can provide a more curated recipe and try to increase the average rating.

### Chart #4:

- Hypothesis: This chart compares the Years a particular recipe has appeared in menus vs the reviews we have for those recipes. This again would help us determine whether the

popular recipes are being reviewed enough. We further believe that the recipes that have been on restaurant menus the longest will have the most reviews ideally.

YearsInRestaurants and ReviewCounts



Quick Note: We have excluded recipes that appeared for less than 100 years on menus to reduce the number of recipes and keep the most popular recipes amongst historical menus.

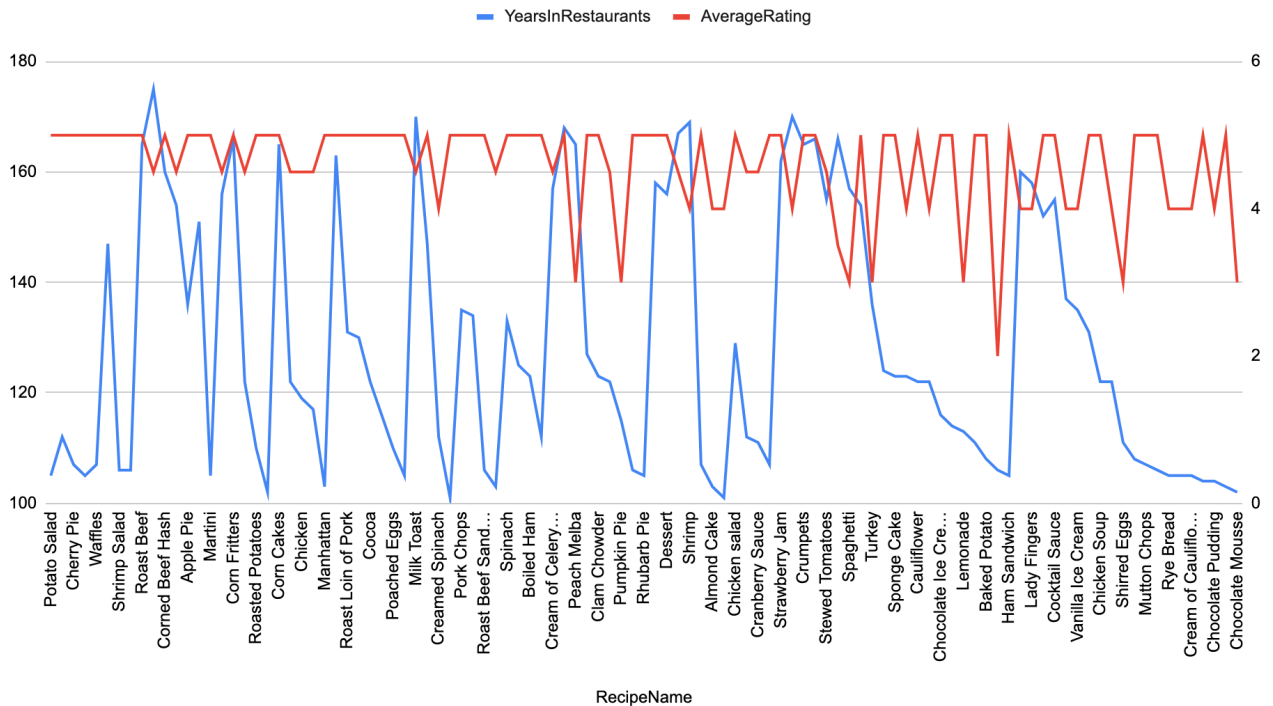
### Interpretation from the Graph:

- This again gives us the same conclusion as Chart #2, we can see that a lot of the 'timeless' recipes do not have enough reviews. This must be ratified to improve the overall quality of CookEasy to increase user retention.

### Chart #5:

- Hypothesis: This chart compares for each recipe the number of years it has been on menus compared to the rating it has on our website. Here as well, our expected result would be that more the number of years it has been on menus, the average rating should be high for that recipe. It will overall help us assess the quality of content on CookEasy.

## YearsInRestaurants and AverageRating



Quick Note: We have excluded recipes that appeared for less than 100 years on menus to reduce the number of recipes and keep the most popular recipes amongst historical menus.

### Interpretation from the graph:

- Here, similar to chart #3, we can verify that for most popular recipes according to the number of years it has been appearing in menus we have a good average rating. There are some recipes which have rating less than 4, which need to be looked at and improved content wise to assure that all the popular ratings and well curated on CookEasy. As mentioned before, this will help us keep user retention and improve quality of CookEasy.