Project Installation and instructions to run

File Summary

- Data scraping code files
  - Website scraping 1: Scrapping.py
  - Website scraping 2: WhatsCooking.py

  -
- Exploratory Data Analysis : EDA.py
- Association Rule Mining: AssociationRuleMining.py
- Recipe Recommendations: Recipe_Recom.py

Installations

Package	Installation Direction

mlxtend	`pip install mlxtend`
	
	`If problem persists, install using conda:`
	`conda install mlxtend --channel conda-forge`

wordcloud	`pip install wordcloud`

Datasets:
1. Instacart: https://www.instacart.com/datasets/grocery-shopping-2017
2. Recipe
   a. Genius Kitchen: https://www.geniuskitchen.com/?ref=nav
   b. United States Department of Agriculture: https://whatscooking.fns.usda.gov/search/recipes

Consolidated Data: File - Datasets_Group4.xlsx

Scrapping :
            Website Source: https://www.geniuskitchen.com/?ref=nav
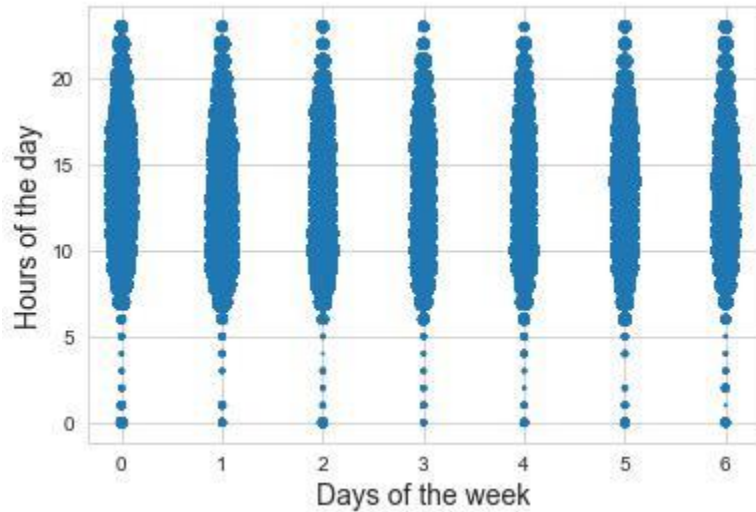Output files   : Recipes.csv
Screenshot   :

```
 RESTART: C:\Users\alibh\OneDrive\Documents\1. Sem2\Python\Project\Scrapping.py
creamy burrito casserole
cracker barrels hashbrowns casserole copycat
tuna casserole
kittencals chicken crescent roll casserole
fantastic taco casserole
best ever tuna noodle casserole
easy breakfast casserole
easy creamy sour cream chicken casserole
kittencals moist cheddar garlic oven fried chicken breast
melt in your mouth chicken breasts
spinach feta stuffed chicken breast quick easy
thai chicken breasts
baked buffalo chicken breasts
chicken breasts in lemon cream sauce
weight watchers mexican chicken breasts
easy bone in split chicken breasts
creamy cajun chicken pasta
ro tel chicken spaghetti
chicken spaghetti
cheesecake factorys louisiana chicken pasta
cajun chicken pasta
tgi fridays copycat bruschetta chicken pasta
carmelas chicken rigatoni copycat romanos macaroni grill
caesar chicken pasta
```
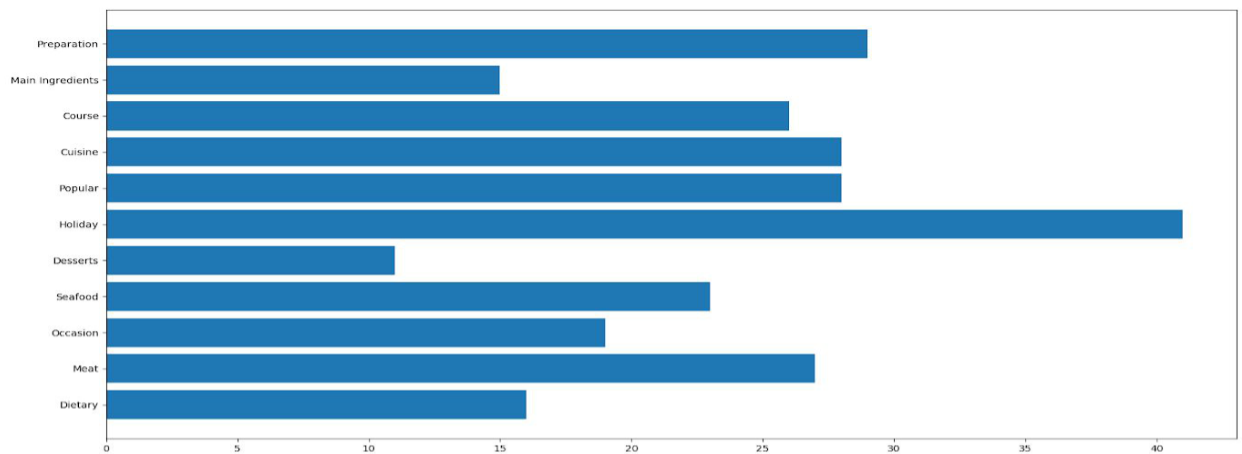
Exploratory Data Analysis:

1. **Department Stocks**: This is a visual showing the different variety of products offered by various departments i.e department-wise product stocks

2. **Popular Purchase Days/Hour**: This is a visual showing days on which number of purchases is more comparatively and also shows how the sales varies throughout the day.

3. **Category Repository**: The following visual shows the distribution of the number of recipes per category and is followed by a Word Cloud plot of the same for better inference.

Association Rule
Screenshot of the results from *Association Rule Mining*
1. Screenshot showing a list of products which are bought more than 5 % of the time and 1% of the time

```
======= RESTART: C:\Users\Vipasha\Homework\AssociationRuleMining.py =======
Products that are bought more than Five percent of the time

   support                 itemsets
0  0.122094  (Bag of Organic Bananas)
1  0.148288                  (Banana)
2  0.064493             (Large Lemon)
3  0.058807          (Organic Avocado)
4  0.078078     (Organic Baby Spinach)
5  0.057580     (Organic Hass Avocado)
6  0.086365     (Organic Strawberries)
7  0.051611             (Strawberries)

Products that are bought more than one percent of the time

     support                                  itemsets  length
0   0.010312              (100% Raw Coconut Water)       1
1   0.018064              (100% Whole Wheat Bread)       1
2   0.012210                  (2% Reduced Fat Milk)       1
3   0.017749              (Apple Honeycrisp Organic)     1
4   0.030841                          (Asparagus)        1
5   0.122094              (Bag of Organic Bananas)       1
6   0.148288                            (Banana)         1
7   0.016700      (Boneless Skinless Chicken Breasts)    1
8   0.023634                    (Broccoli Crown)         1
9   0.015399                  (Bunched Cilantro)         1
10  0.019889                          (Carrots)          1
11  0.010700                      (Clementines)          1
12  0.011644                  (Clementines, Bag)         1
13  0.019050                    (Cucumber Kirby)         1
14  0.016480              (Extra Virgin Olive Oil)       1
15  0.011466              (Feta Cheese Crumbles)         1
16  0.021127                  (Fresh Cauliflower)        1
17  0.010102                  (Fresh Ginger Root)        1
18  0.015431                            (Garlic)         1
19  0.011350                (Granny Smith Apples)        1
20  0.012053        (Grape White/Green Seedless)         1
21  0.013658                  (Grated Parmesan)          1
22  0.020015                (Green Bell Pepper)          1
23  0.011413                    (Green Onions)           1
24  0.019438                      (Half & Half)          1
25  0.012987                    (Hass Avocado)           1
26  0.018190                  (Hass Avocados)            1
27  0.028491                (Honeycrisp Apple)           1
28  0.015462                (Jalapeno Peppers)           1
29  0.015232              (Large Alfresco Eggs)          1
..       ...                               ...         ...
91  0.011508              (Sparkling Lemon Water)        1
92  0.014371      (Sparkling Natural Mineral Water)      1
93  0.027117        (Sparkling Water Grapefruit)         1
94  0.017497                      (Spring Water)         1
```

2. Screenshot showing a list of products which are bought together more than 20 % of the time

```
102  0.030128                                  (Yellow Onions)          1
103  0.017602    (Bag of Organic Bananas, Organic Baby Spinach)         2
104  0.010154       (Organic Cucumber, Bag of Organic Bananas)          2
105  0.018882     (Bag of Organic Bananas, Organic Hass Avocado)        2
106  0.014204      (Organic Raspberries, Bag of Organic Bananas)        2
107  0.024316    (Bag of Organic Bananas, Organic Strawberries)         2
108  0.017046                            (Banana, Large Lemon)           2
109  0.010574                               (Limes, Banana)              2
110  0.017760                      (Organic Avocado, Banana)             2
111  0.015882              (Banana, Organic Baby Spinach)                2
112  0.017099              (Banana, Organic Strawberries)                2
113  0.015389                      (Strawberries, Banana)                2
114  0.013039                       (Limes, Large Lemon)                 2
115  0.010574              (Organic Avocado, Large Lemon)                2
116  0.011046        (Organic Avocado, Organic Baby Spinach)             2
117  0.012924    (Organic Baby Spinach, Organic Strawberries)           2
118  0.010081     (Organic Strawberries, Organic Blueberries)           2
119  0.012210    (Organic Hass Avocado, Organic Strawberries)           2
120  0.013228    (Organic Raspberries, Organic Strawberries)            2

[121 rows x 3 columns]
Products that are bought more than 20 percent of the times together

           antecedents            consequents  antecedent support  ...     lift  leverage  conviction
0   (Organic Baby Spinach)  (Bag of Organic Bananas)         0.078078  ...  1.846498  0.008070    1.133435
1      (Organic Cucumber)   (Bag of Organic Bananas)         0.036831  ...  2.258132  0.005658    1.212083
2   (Organic Hass Avocado)  (Bag of Organic Bananas)         0.057580  ...  2.685864  0.011852    1.306269
3    (Organic Raspberries)  (Bag of Organic Bananas)         0.044058  ...  2.640427  0.008824    1.295574
4   (Organic Strawberries)  (Bag of Organic Bananas)         0.086365  ...  2.306004  0.013771    1.221944
5         (Large Lemon)                 (Banana)         0.064493  ...  1.782434  0.007483    1.157711
6             (Limes)                 (Banana)         0.048349  ...  1.474850  0.003404    1.090125
7      (Organic Avocado)             (Banana)         0.058807  ...  2.036563  0.009039    1.220214
8   (Organic Baby Spinach)            (Banana)         0.078078  ...  1.371740  0.004304    1.069201
9         (Strawberries)             (Banana)         0.051611  ...  2.010754  0.007736    1.213560
10            (Limes)            (Large Lemon)         0.048349  ...  4.181682  0.009921    1.280972
11        (Large Lemon)               (Limes)         0.064493  ...  4.181682  0.009921    1.192814
12   (Organic Blueberries)  (Organic Strawberries)       0.039275  ...  2.972008  0.006689    1.229123
13  (Organic Hass Avocado)  (Organic Strawberries)       0.057580  ...  2.455399  0.007238    1.159524
14   (Organic Raspberries)  (Organic Strawberries)       0.044058  ...  3.476387  0.009423    1.305637

[15 rows x 9 columns]

Product which are more than 2.5 times likely to be bought together

           antecedents            consequents  antecedent support  ...     lift  leverage  conviction
2   (Organic Hass Avocado)  (Bag of Organic Bananas)         0.057580  ...  2.685864  0.011852    1.306269
3    (Organic Raspberries)  (Bag of Organic Bananas)         0.044058  ...  2.640427  0.008824    1.295574
10            (Limes)            (Large Lemon)         0.048349  ...  4.181682  0.009921    1.280972
11        (Large Lemon)               (Limes)         0.064493  ...  4.181682  0.009921    1.192814
12   (Organic Blueberries)  (Organic Strawberries)       0.039275  ...  2.972008  0.006689    1.229123
14   (Organic Raspberries)  (Organic Strawberries)       0.044058  ...  3.476387  0.009423    1.305637

[6 rows x 9 columns]
```

Recipe Suggestion:

Screenshot of the results of *Recipe Recommendation* for a given order

	RecipeName	Ingredients	Directions	CookingTime	Serves
635	bacon wrapped pork chops with bbq sauce	4 boneless pork chops (about 1 inch thick);1/...	Preheat oven to 400Â°F.;On work surface layer ...	READYIN:1hr5mins	SERVES:4
646	mardi gras coleslaw	2 cups shredded red cabbage;2 cups shredded ...	Toss both kinds of cabbage, the grated carrots...	READYIN:15mins	SERVES:5-8
651	blt barbecue chicken salad	1/2 cup mayonnaise;3 -4 tablespoons barbecu...	In a small bowl, mix the mayo, bbq sauce, onio...	READYIN:10mins	SERVES:8

Complete set of instructions to run the project:

1. Install the required packages as directed in the installation section table
2. Run EDA.py for scrapping and visualization
3. Open DeliciousDaily.py in Idle which calls all the other required modules
4. The first module to run is Scrapping.py which scraps the data and creates text files for convenience so that the scraping code need not be run after the first time and the text files can be used instead.
5. After scraping, the data has been consolidated in a single excel containing separate excel sheets corresponding to different information
6. Next, Exploratory Data Analysis was performed and its results are shown above.
7. After EDA, Association Rule Mining was performed on the Instacart dataset and discovered association between different types products that are bought together
8. The final step was recommending recipes based on a customer's cart contents. This application recommends the top 3 recipes fitting the customer's purchase order.