Methodology and Experiment

Effect of dietary patterns on the cause (ACR) as well as the mortality of Chronic Kidney Disease (CKD) patients

Data Analytics: Major Research Project

Sayed Ahmed
MSc in Data Science and Analytics
Ryerson University
500869723

Supervisor
Dr. Youcef Derbal
Information Technology Department
Ted Rogers School of ITM
Ryerson University

Table of Contents

Methodology

In this section, the methodology utilized for the research will be provided. A diagram showing all the steps in the methodology is provided in the next page.

The primary purpose of this research is to assess the effect of dietary patterns on the cause (ACR - Albumin Creatinine Ratio) as well as the mortality and survival in relation to chronic kidney disease (CKD)/End Stage Renal Disease (ESRD). A dataset released by CDC/Health.gov with the dietary habits and a CKD measure named Albumin Creatinine Ratio (ACR) of around 10,000 individuals are studied. Also, age group based mortality and survival of CKD/ESRD patients provided by USRDS are studied. Afterwards, utilized Principal Component Analysis to identify the most important food groups and subgroups affecting the ACR value and the Mortality/Survival then utilized Statistical Regression and Factor analysis to understand the correlation of ACR, and Mortality/Survival to dietary patterns. Machine learning approaches such as Regression, Polynomial Regression, and Bayesian with or without 10 fold cross validations are applied on the datasets to understand if dietary patterns can be used to predict ACR values and mortality. The experiments showed 57 to 95% accuracy in the test data depending on the methodology applied.

Additionally, dietary recommendations as provided by health.gov are utilized for food groups/subgroups. Association of CKD mortality and ACR values with deviation from the recommended amount is studied. *Another study [11] on shifting from current recommendations conducted by health.gov is explored. Will the recommended shift [11] from current diet style [15] can have an improved outcome or not is also explored (optional goal).*

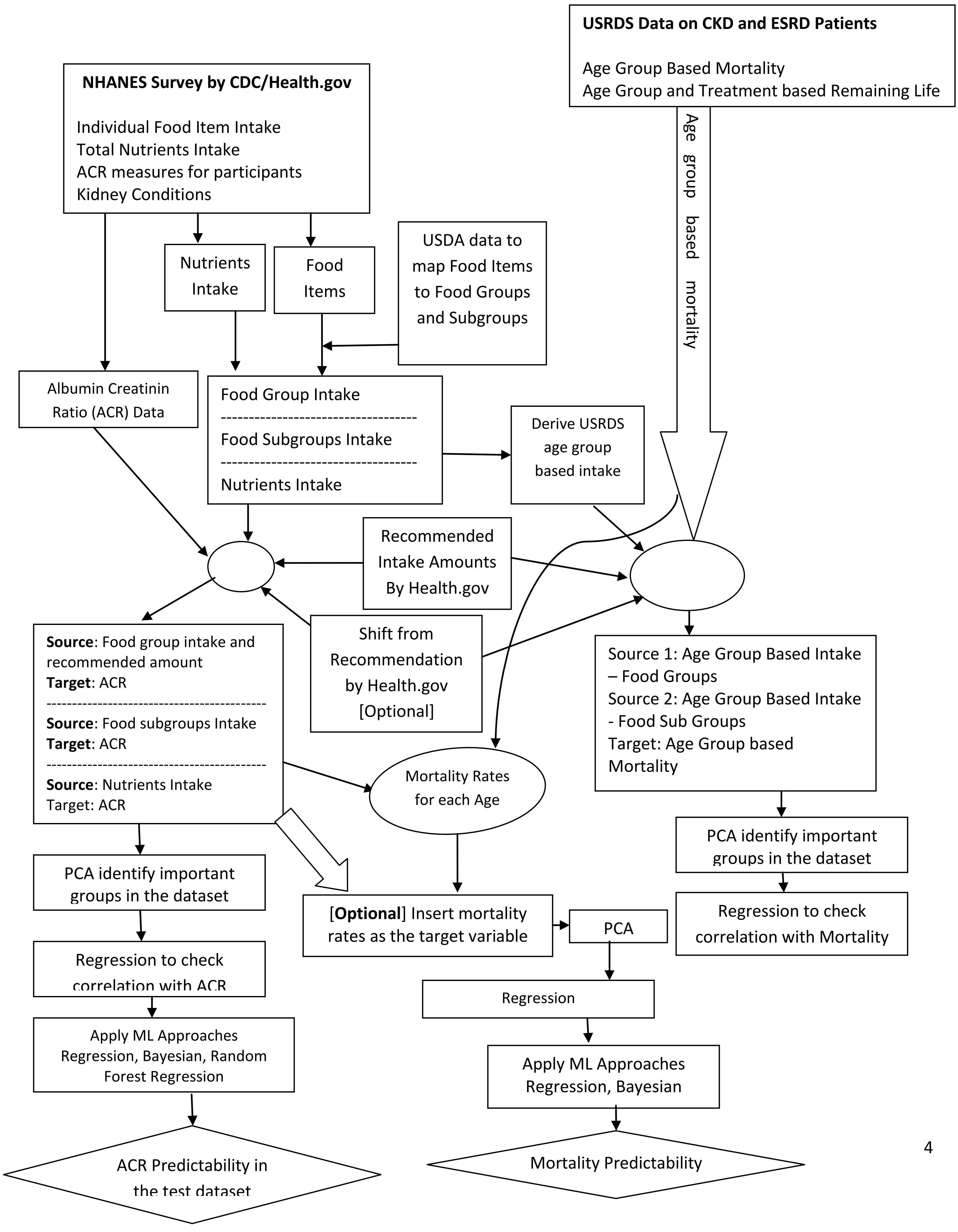
Study Selection

For dietary patterns, CKD measures (such as Albumin Creatinine Ratio - ACR) , and Kidney condition measures, a dataset from the **National Health and Nutrition Examination Survey** on dietary habits conducted by the Centers for Disease Control and Prevention (CDC) [10] was used. The survey has data from 1996 to 2016 [10]. This study primarily utilized data for 2015-2016. The survey recorded 24 hours individual food item intake amount. Two surveys were taken within 3 to 10 days apart. Each survey provided food item intake amount in a day, also mentioned the diet style, and diet-restrictions. Individual food items are represented using USDA food code. The survey also provided total nutrients data. CDC also released examination, laboratory, demographics, and other related data for those participants.

For mortality and survival information, dataset from the United States Renal Data System (USRDS) on CKD and ESRD [16, 17] was utilized. “USRDS investigates the transition of care from CKD to ESRD and end-of-life care for those with advanced kidney disease” [19]. USRDS also releases data on the Incidence, Prevalence, Patient Characteristics, and Treatment Modalities on CKD, and ESRD patients. USRDS reports the survival and mortality using metrics such as Mortality rates: ESRD patients, Mortality rates: Dialysis patients, Total Mortality Count, 90 day survival for dialysis and/or transplantation patients, 10 year survival for dialysis and/or transplantation patients, Avg. Expected remaining lifetime with or without pre-condition and treatment options used. The data are either aggregated or patient specific detail data. However, only aggregated data are public where patient specific data access requires special request and permission. This research utilized only the public dataset.

The dietary survey data (NHANES) represented the food items taken by the participants using USDA food codes [14, 12, 13]. Hence, USDA food codes [14, 12, 13] are used to assign food groups and subgroups to the NHANES [10] survey data to properly group/subgroup the dietary intake of the participants with some customizations as provided in the appendix.

Figure: Methodology in a Diagram



Data Synthesis

NHANES survey data as provided for two days are averaged to get the intake amount for one day. Both individual food item data and nutrients intake data are averaged. USDA food codes are used to map food items to food groups and subgroups.

ACR and Kidney condition data for each individual are merged with the averaged food group/subgroup and nutrients data. ACR values are used as the target variable for couple of experiments and study. With this data food group recommendations from health.gov were also merged. This dataset was used for ACR association study.

For one mortality/survival study mortality rates from USRDS for each age were merged with the above data. Mortality/survival is used as the target variable.

For both of the above cases, Principal Component Analysis (PCA) was applied to find out important food groups and subgroups. Afterwards, Regression was applied to find association with ACR and Mortality. Afterwards, Machine Learning approaches such as Linear Regression, Polynomial Regression, Random Forest Regression, Bayesian prediction with or without 10 fold cross validations were applied to study the predictability of ACR Values and Mortality in the test dataset. Test dataset was also part of the above datasets.

For another mortality/survival study, the above synthesized datasets were aggregated for USRDS age groups to calculate average food group/subgroup intake by age groups. With that aggregation mortality/survival data were merged. Mortality/survival is used as the target variable. For this, PCA and Regression were used to find association between food groups and CKD/ESRD mortality

Experiment and outcome

Experiment Overview

Experiments as provided below are planned to achieve the goal of the research. Only first six sets of experiments will be conducted where the 7th set will be an option. Others (set 8 and 9) are not conducted; however, can provide valuable insight to justify the goal.

Set 1: Mortality and CKD: Food Groups **Primary Input Dataset**: NHANES survey aggregated to calculate average food item intake by USRDS like age groups **Target Variable: ESRD:** Avg. Annual Mortality rates
Experiment 1.1: Identify contributing and important food groups in the dataset using PCA a) Using Actual Intake Amount b) Using ratios of intake and recommended high
Experiment 1.2: Find out correlation (using Pearson's correlation and regression) between CKD mortality and important food groups as found using PCA in experiment 1. a) Using Actual Intake Amount b) Using ratios of intake amount and recommended high
Set 2: Mortality and CKD: Food Sub Groups **Primary Input Dataset**: NHANES survey aggregated to calculate average food item intake by USRDS like age groups **Target Variable: ESRD:** Avg. Annual Mortality rates
Experiment 2.1: Identify important food sub groups in the dataset using PCA a) Actual Intake Amount
Experiment 2.2: Similar to experiment 1.2 (Regression); however, used food subgroups and actual intake only
Set 3: ACR and Food Groups **Primary Input Dataset**: NHANES survey data for each participant averaged for two surveys **Target Variable:** Albumin Creatinine Ratio (ACR)
Experiment 3.1: Identify contributing food groups in the input dataset using PCA. This is different than Experiment 1 because entire survey is being used here; not the aggregated data by age groups

Experiment 3.2: Find out correlation (using Pearson's correlation, and regression) between ACR Values and important food groups as found using PCA in experiment 3.1.
Set 4: ACR Values and Nutrients Utilize the same experiments as done for ACR and Food Groups. However, nutrients intake with or without combining with food groups 4.1 PCA to identify contributing factors 4.2 Regression to find correlations among factors found in experiment 4.1
Set 5: ACR Values and Food Subgroups **Primary Input Dataset**: NHANES survey data for each participant averaged for two surveys **Target Variable:** Albumin Creatinine Ratio (ACR) Similar experiments like set 3 and set 4. However, use food subgroups as the input/source variables
Set 6: Experiments using Regression: ACR Values and Food Subgroups
And then utilize Machine Learning Approaches for Mortality Prediction on Test Dataset. Input dataset from Set 6 can be used here as the Input dataset
Experiment 6.1: ACR value prediction using linear regression. (ACR class can also be an option). **Experiment 6.2** Use 10 folds cross validations where possible. **Goal:** Check the % of predictability in the test dataset. Precision, recall, or similar might be calculated
Experiment 6.3: Conduct experiment 6.1; however use Polynomial Regression **Experiment 6.4:** With 10 Folds Cross Validations
Experiment 6.5: Conduct experiment 6.1; however, use Random Forest Regression with or without 10 Folds Cross Validations. Utilize Polynomial Regression in the process.
Experiment 6.6: Conduct experiment 6.1; however, use Bayesian prediction with or without 10 Folds Cross Validations. b) Use Polynomial Fit
Set 7: CKD Mortality using Survey data i.e. No aggregation on Age Groups Input Data: Bring CKD mortality data to each participant using the corresponding age Use PCA (to find contributing food groups and subgroups) and then Regression to find correlation between mortality and Food Groups/Subgroups/ACR values
Set 8: Experiments using Regression: No aggregated (on age groups) survey data And then utilize Machine Learning Approaches for Mortality Prediction on Test Dataset. Input dataset from Set 7 can be used here as the Input dataset
Set 9: Remaining Life for CKD Patients and Food Groups/Subgroups: No aggregated survey data Input Data: Bring remaining life data to each participant using the corresponding age and CKD status Use PCA (to find contributing food groups and subgroups) and then Regression to find correlation between remaining life and Food Groups/Subgroups/(ACR values optional) Will not work on this as data are not sufficient and the project has becoming long.

Experiment Details and Outcome:

Set 1: Mortality and CKD: Food Groups Only

Experiment: Identify contributing and important food groups in the dataset using PCA

Outcome (Plots are given below):

- First three components explain the data by 91%
- **Contributors to the First Component**: Sugar, Oil, Protein and then Vegetables **negatively** (-0.6 approx).

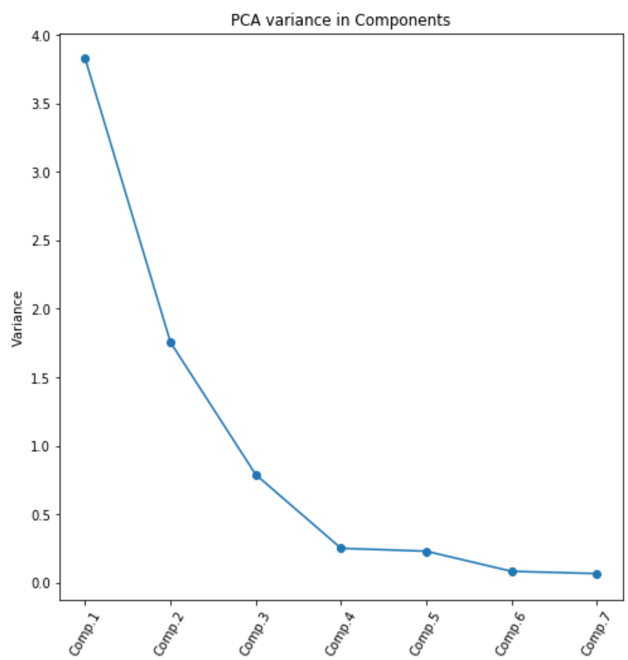
Dairy somewhat (0.5) **positively**, and Fruits (.3) minimally **positively**

- **Contributors to the 2nd Component**: grain strongly and **negatively**. Oil and protein minimally and **negatively**. Vegetables strongly and **positively**. Sugars minimally positively
- **Contributors to the 3rd component**: Fruits **positively** where sugars minimally positive.
- **Contributors for 4th component**: Oil strongly and **negatively**

Considering all components, Grain and Oil contribute the most and negatively then protein and sugar. Fruits contributes the most positively then vegetable and dairy.

The above outcome (also the numbers) is primarily taken from experiment using intake amount to high end of recommendation amount. Analysis with actual intake amounts shows very similar outcome as above. Regressions will be done with all food groups as all are found to be contributing.

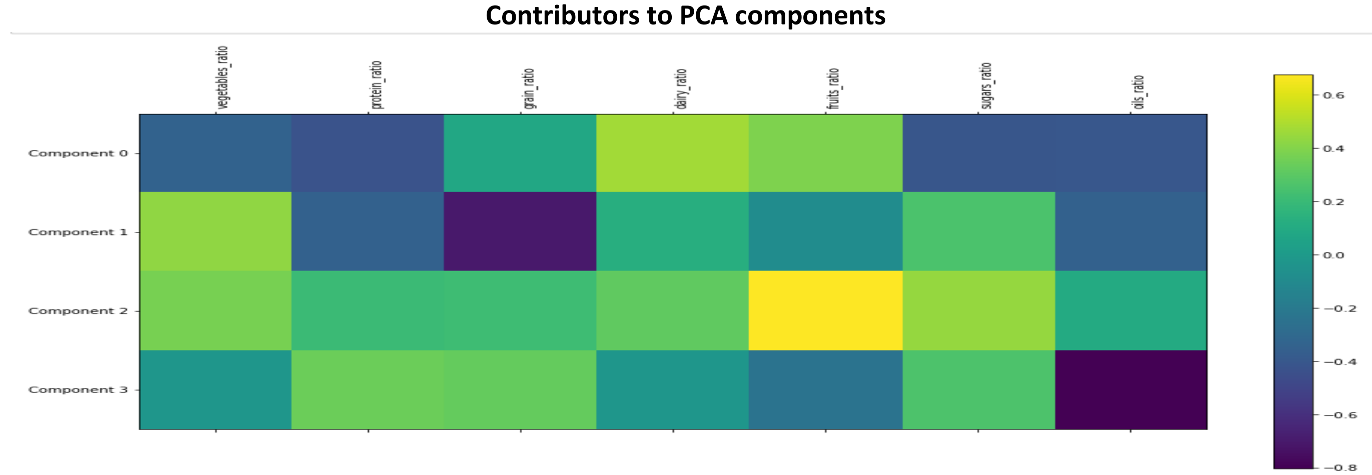
Fig below: PCA components, and contributors to PCA components



Importance of components:

	sdev Standard deviation	varprop Proportion of Variance	cumprop Cumulative Proportion
PC1	1.956073	0.546603	0.546603
PC2	1.324680	0.250683	0.797286
PC3	0.888272	0.112718	0.910004
PC4	0.501019	0.035860	0.945864
PC5	0.479280	0.032816	0.978680

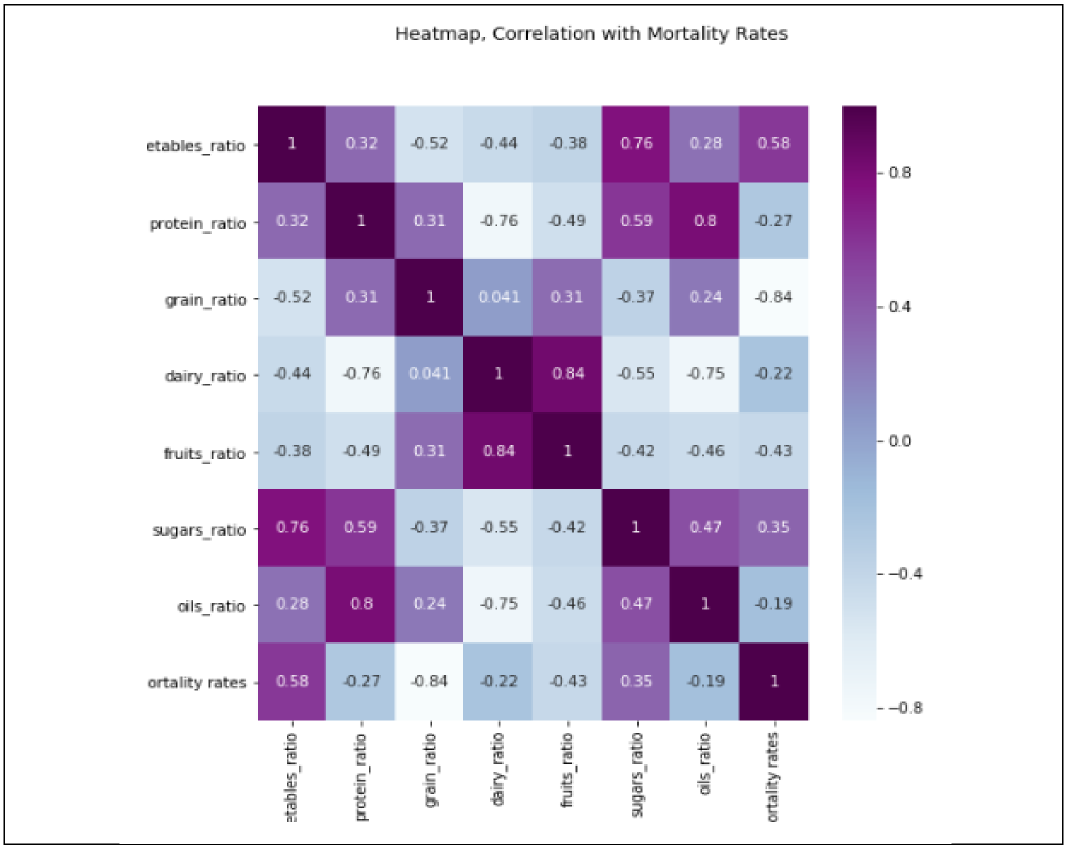
Contributors to PCA components



Experiment: Find out correlation (using Pearson's correlation and regression) between CKD mortality and important food groups as found using PCA in the previous experiment.

Regression:
Grain (-0.84) and Fruits (-0.43) respectively show strong negative correlation with Mortality rates i.e. if Grain and protein taken less, mortality is high. Vegetables show positive (0.58) correlation where sugars slightly positive (0.35) correlation i.e. when vegetables are taken high, mortality is also high. However, data shows this in older adults. Still ratios were used, age might have a bias. This does not show conformity with doctors' recommendation on taking more vegetables for CKD patients. If this analysis has to be right: moderate (not high) vegetable intake might be an option for older adults.

The above results are from ratios (experiment 2.2). However, PCA and Regression using Actual Intake Amounts (Exp 2.1) show similar outcome where sugar (0.0004) shows virtually no effect. Hence, this research will conclude no effect on mortality for added sugars.



Set 2: Mortality and CKD: Food Sub Groups

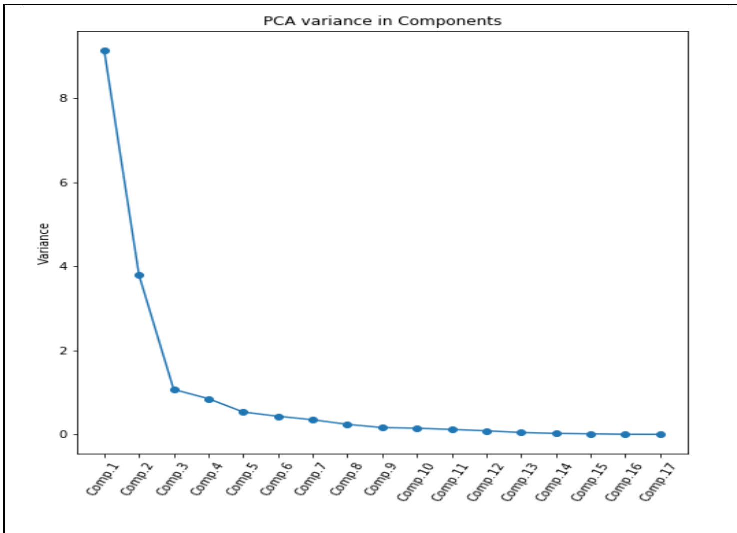
Experiment: Identify important food sub groups in the dataset using PCA a) Actual Intake Amount only

Outcome:
- First five components explain the data by 90%
- **Contributors for the First Component:**
  - **Positively** (0.4): Milks and milk drinks
  - **Negatively** (-0.3 to -0.4): Dark-green vegetables, Starchy vegetables, Whole grains, Meat Poultry and Eggs, Oils, Water noncarbonated intake, Nonalcoholic beverages taken
- **Contributors for 2nd Component:**
  - **Positively**: Red and orange vegetables, Other vegetables
  - **Negatively**: Nuts Seeds and Soy Products, Added Sugars/Sugars and sweets, Alcoholic beverages
- **Contributors for 3rd component:**

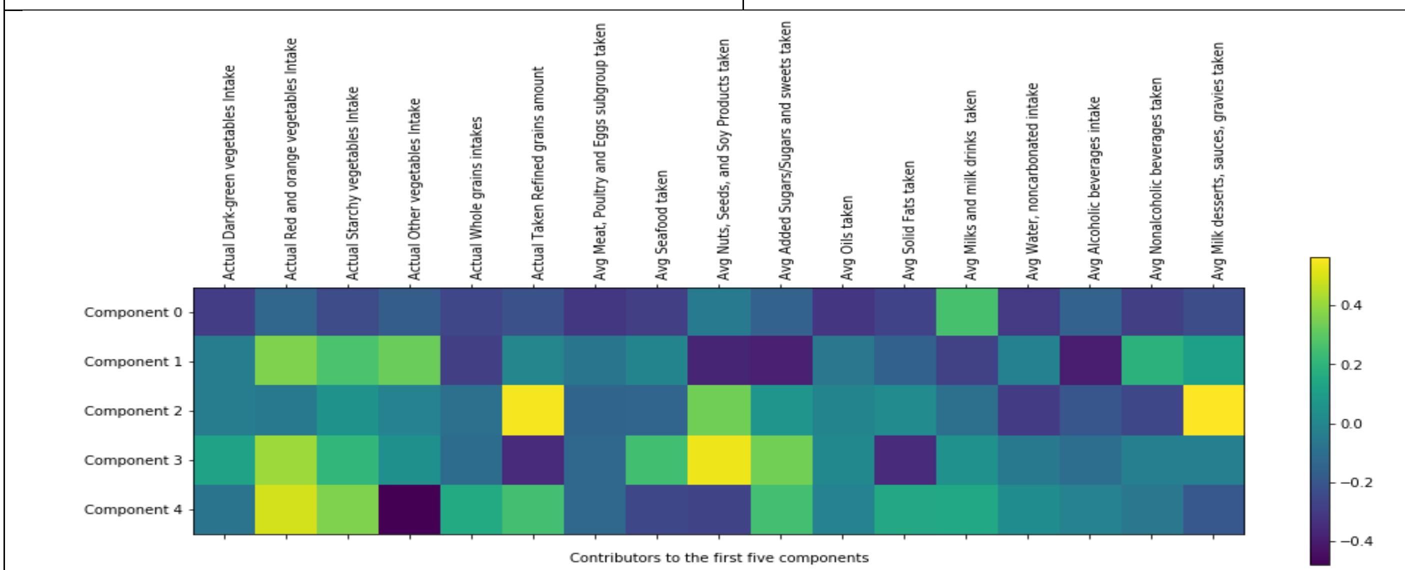
   - **Positively**: Refined grains, Milk desserts sauces gravies
   - **Negatively**: Water noncarbonated
- **Features that do not contribute significantly:** Seafood, Solid Fats
- **Considering all Five Components:**
   - **Most Positively** : Refined grains, Milk desserts sauces gravies, Nuts Seeds and Soy Products, Red and orange vegetables
   - **Most Negatively:** Other vegetables, Added Sugars/Sugars and sweets, Alcoholic beverages

Regression analysis will be done with all food sub groups except Seafood, Solid Fats

Fig Below: PCA components and Contributors to the PCA components



	sdev	varprop	cumprop
PC1	3.021925	5.371782e-01	0.537178
PC2	1.948069	2.232338e-01	0.760412
PC3	1.034469	6.294863e-02	0.823361
PC4	0.921359	4.993539e-02	0.873296
PC5	0.730708	3.140792e-02	0.904704
PC6	0.657517	2.543113e-02	0.930135

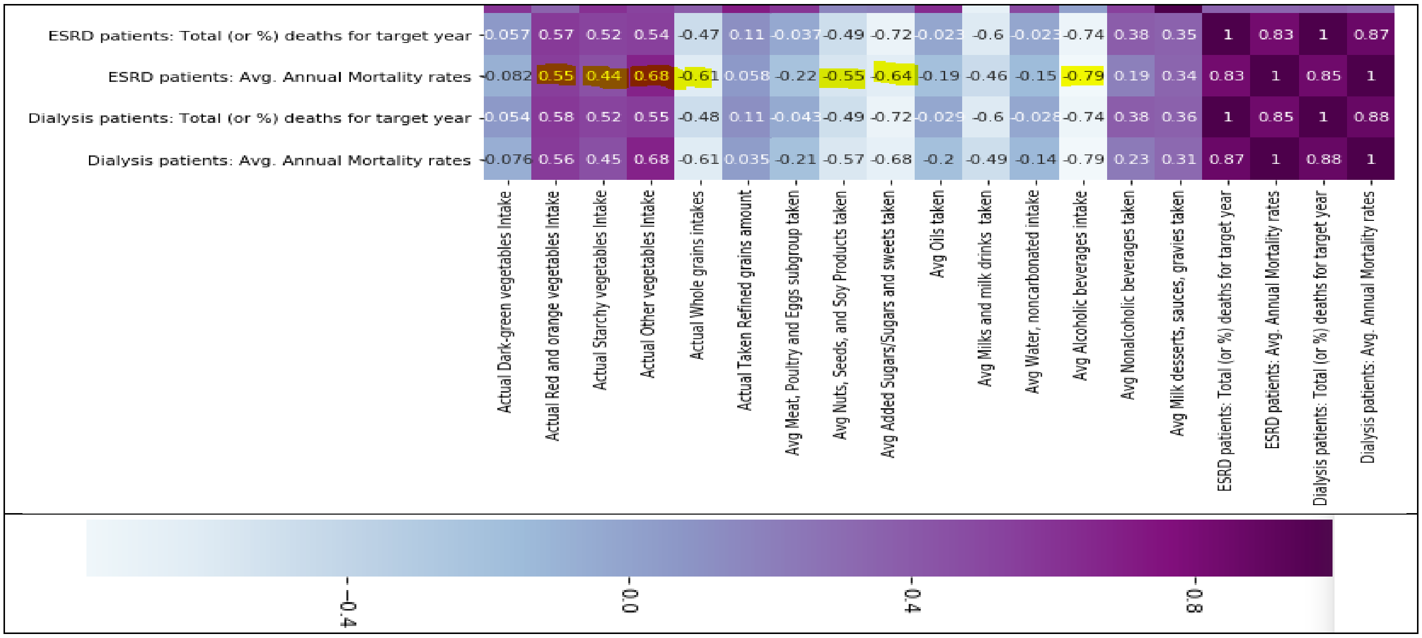


Experiment: Use Regression to find correlations to Mortality rates with food subgroups. On Actual Intake Amount

Outcome:

- **Most Positively Correlated**: Other vegetables, Red and orange vegetables, Starchy vegetables,
- **Most Negatively Correlated**: Alcoholic beverages (-0.79), Added Sugars/Sugars and sweets (-0.64), Whole grains (-0.61), Nuts, Seeds, and Soy Products (-0.55)

Corresponding line chart indicating positive and negative correlations are provided in the appendix.

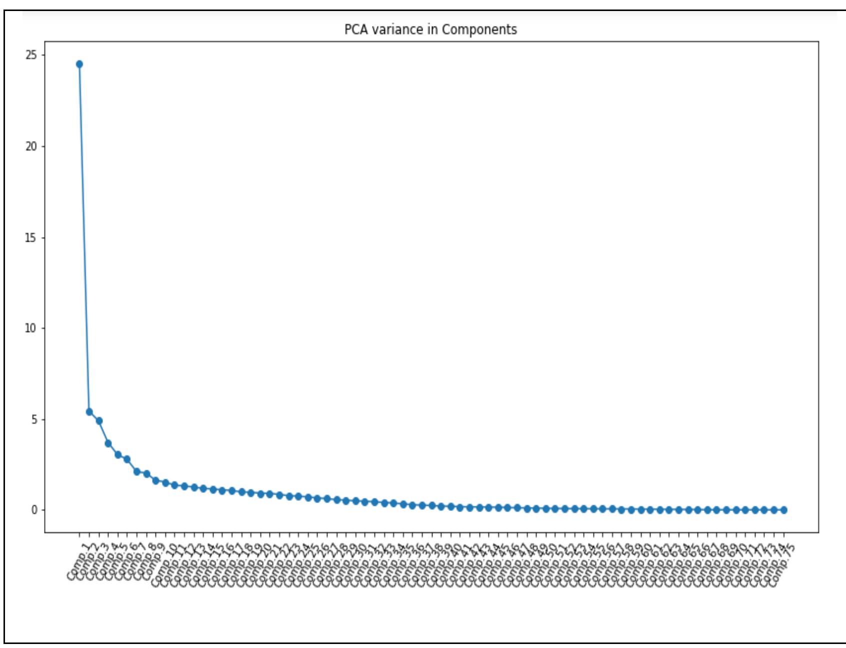


Set 3 and Set 4: ACR Values with Food Groups and Nutrients Intake separately or in combination

Notes on input Data: Combined food group based intake data and nutrients intake data into the same dataset then applied PCA and Regression together (food groups and nutrients) and separately.

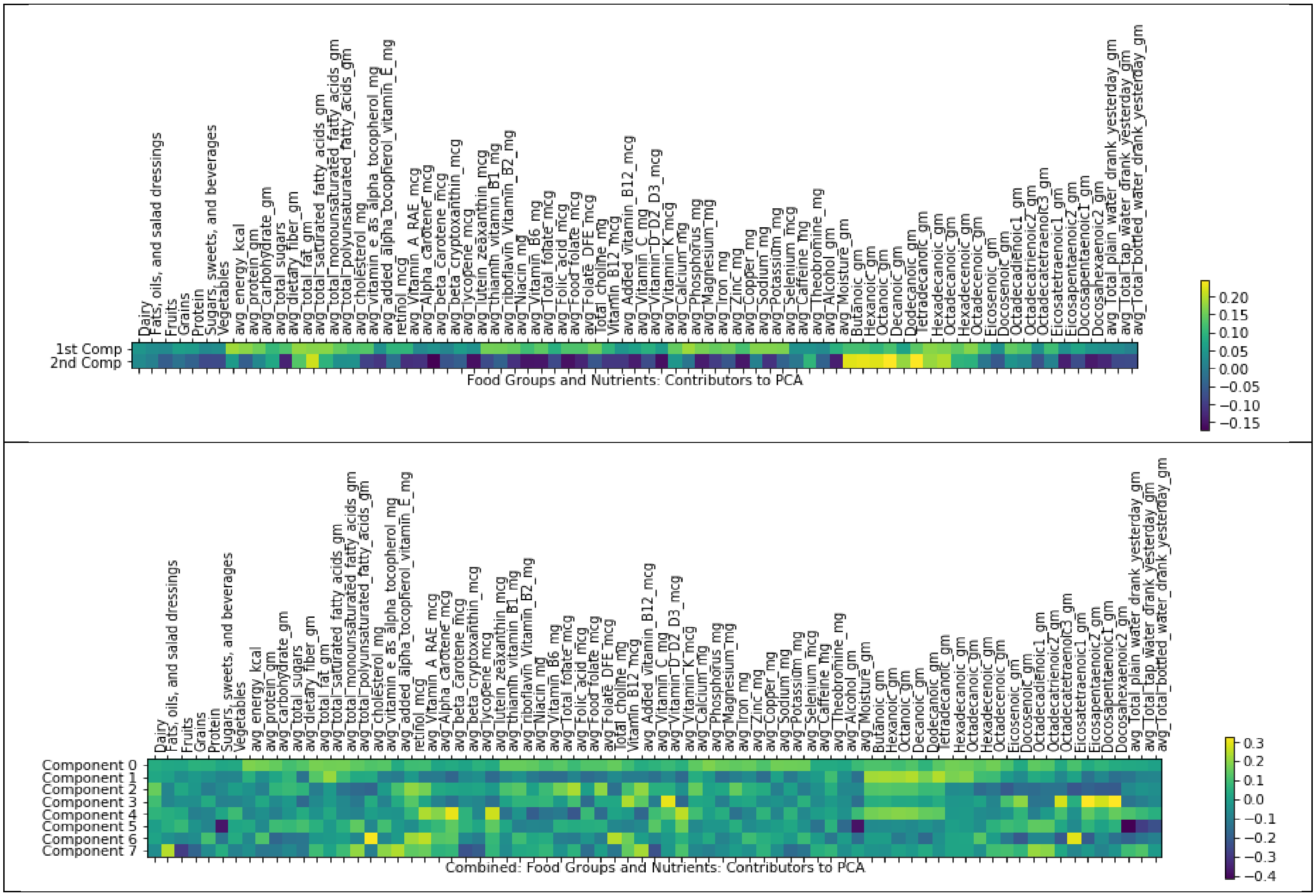
Experiment with PCA to identify contributing factors:

PCA: Combined Food Group and Nutrients



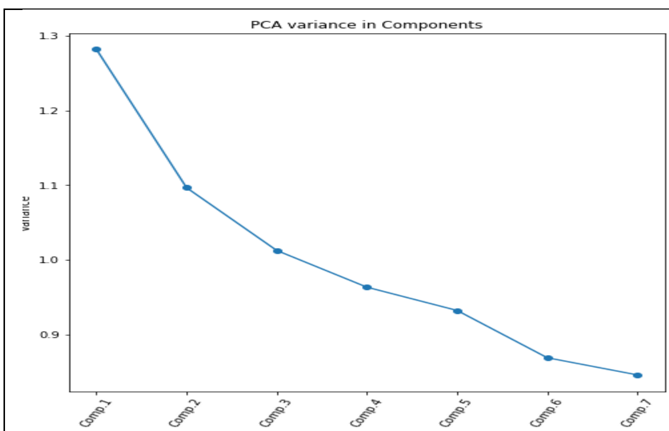
Importance of components:

	sdev	varprop	cumprop
PC1	4.922869	0.356392	0.356392
PC2	2.315556	0.078850	0.435242
PC3	2.170895	0.069306	0.504547
PC4	1.896750	0.052907	0.557454
PC5	1.722163	0.043615	0.601070
PC6	1.564879	0.036012	0.637082
PC7	1.429578	0.030054	0.667136
PC8	1.336302	0.026260	0.693397
PC9	1.224046	0.022034	0.715430
PC10	1.201172	0.021218	0.736648
PC11	1.155583	0.019638	0.756286
PC12	1.123579	0.018565	0.774851
PC13	1.068706	0.016796	0.791647
PC14	1.046879	0.016117	0.807764
PC15	1.023992	0.015420	0.823184



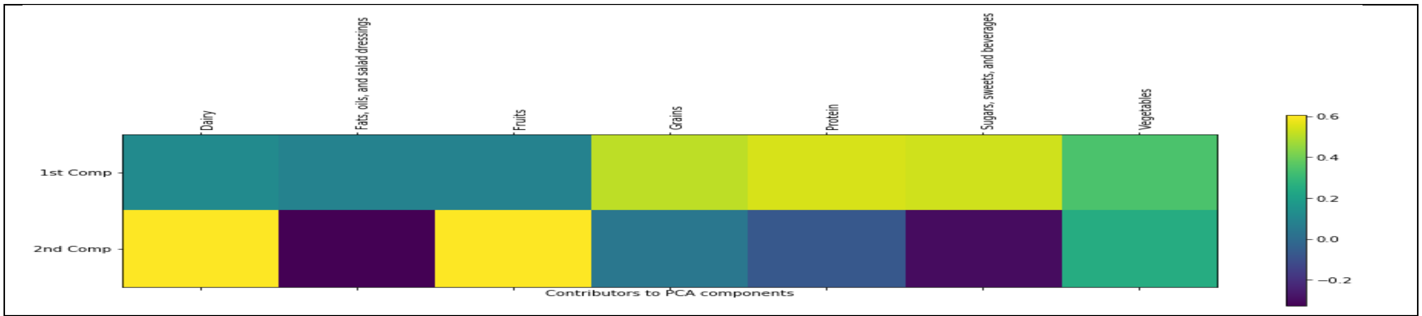
PCA Outcome, Contributing food groups and nutrients: Dairy, 'Fats, oils, and salad dressings', Fruits, Grains, Protein, 'Sugars, sweets, and beverages', Vegetables, Avg_energy_kcal, avg_protein_gm, avg_carbohydrate_gm, avg_total_fat_gm, avg_total_saturated_fatty_acids_gm, avg_total_monounsaturated_fatty_acids_gm, avg_total_polyunsaturated_fatty_acids_gm, avg_lutein_zeaxanthin_mcg, avg_thiamin_vitamin_B1_mg, avg_riboflavin_Vitamin_B2_mg, avg_Niacin_mg, avg_Calcium_mg, avg_Phosphorus_mg, avg_Magnesium_mg, avg_Iron_mg, avg_Zinc_mg, avg_Copper_mg, avg_Sodium_mg, avg_Potassium_mg, avg_Selenium_mcg, Hexadecenoic_gm, Octadecenoic_gm

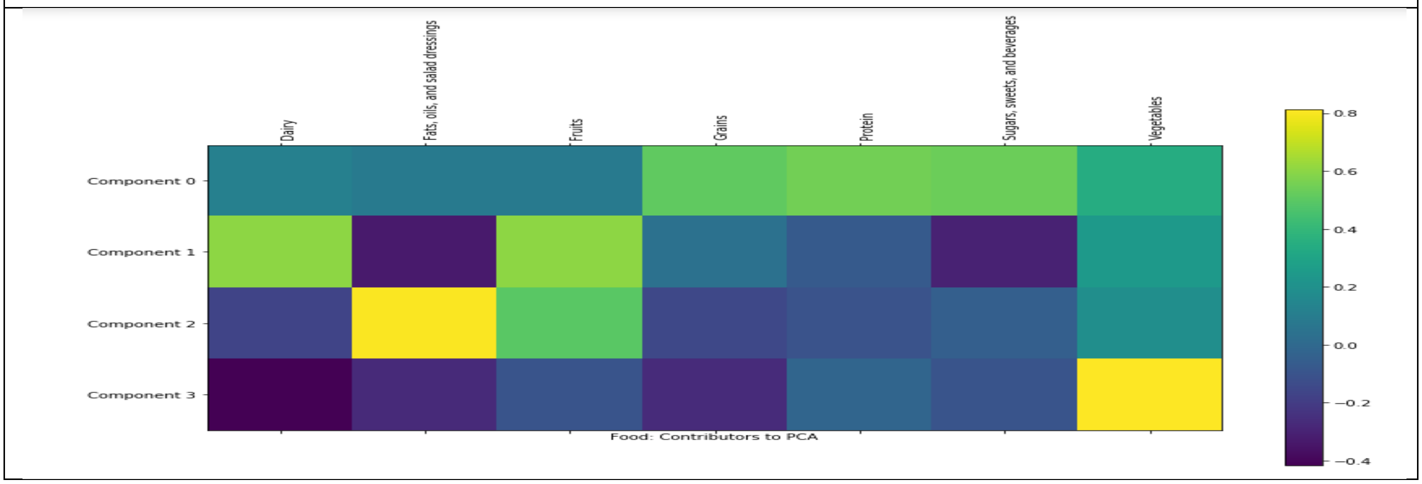
PCA: Food Group Only:



Importance of components:

	sdev	varprop	cumprop
PC1	1.132112	0.183097	0.183097
PC2	1.047078	0.156625	0.339721
PC3	1.006152	0.144620	0.484341
PC4	0.981421	0.137598	0.621940
PC5	0.965417	0.133147	0.755087
PC6	0.932037	0.124099	0.879186
PC7	0.919619	0.120814	1.000000

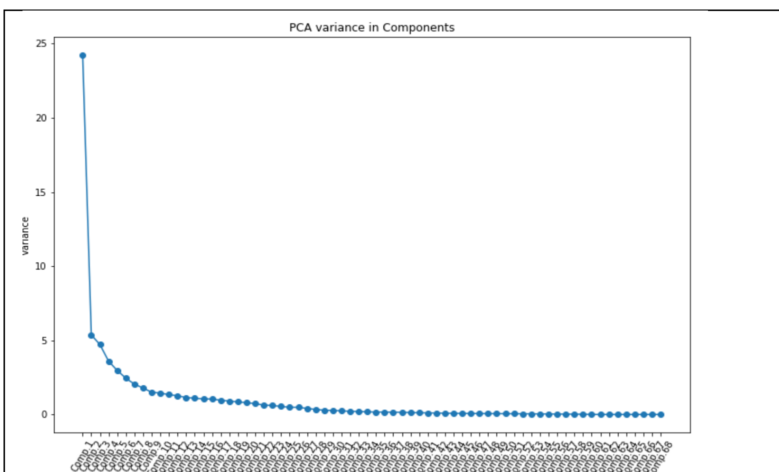




Notes on Food Groups:

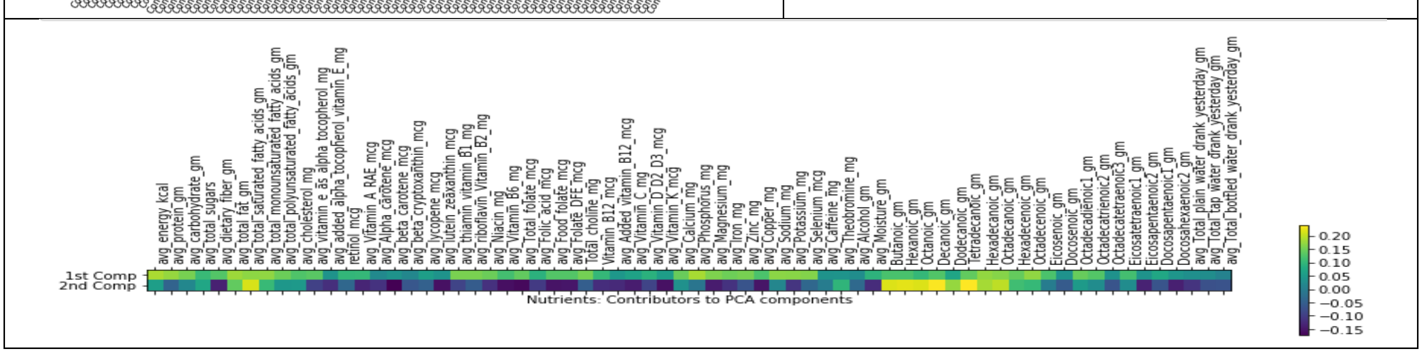
Grain, Protein, and Sugars were important in the first component. Dairy and Fruits were important in the second component. Fats in the third component then fruits were important. Vegetables in the 4th component

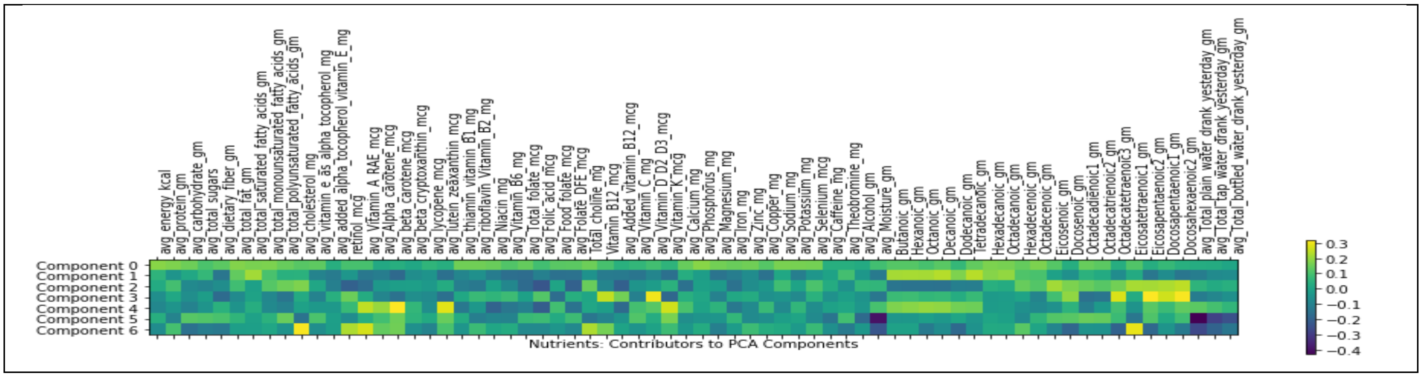
PCA: Nutrients:



Importance of Components

	sdev	varprop	cumprop
PC1	4.922869	0.356392	0.356392
PC2	2.315556	0.078850	0.435242
PC3	2.170895	0.069306	0.504547
PC4	1.896750	0.052907	0.557454
PC5	1.722163	0.043615	0.601070
PC6	1.564879	0.036012	0.637082
PC7	1.429578	0.030054	0.667136
PC8	1.336302	0.026260	0.693397
PC9	1.224046	0.022034	0.715430
PC10	1.201172	0.021218	0.736648
PC11	1.155583	0.019638	0.756286
PC12	1.123579	0.018565	0.774851
PC13	1.068706	0.016796	0.791647
PC14	1.046879	0.016117	0.807764
PC15	1.023992	0.015420	0.823184



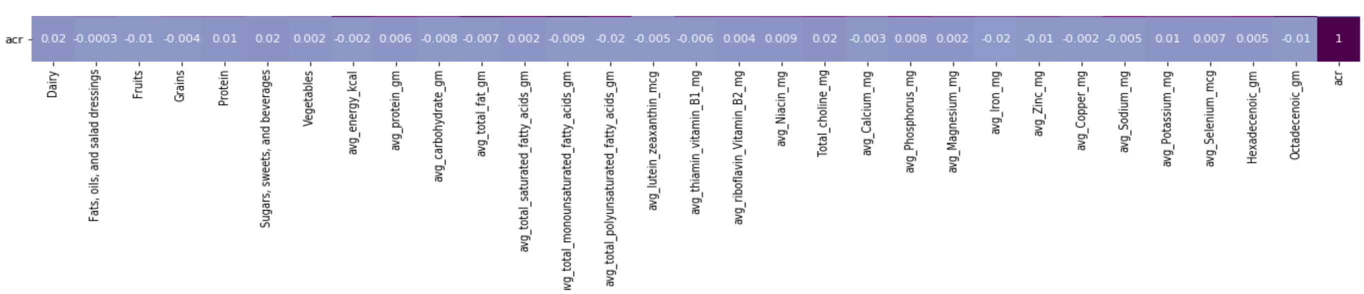


Experiment using Regression to find correlation between ACR and Food Groups/Nutrients

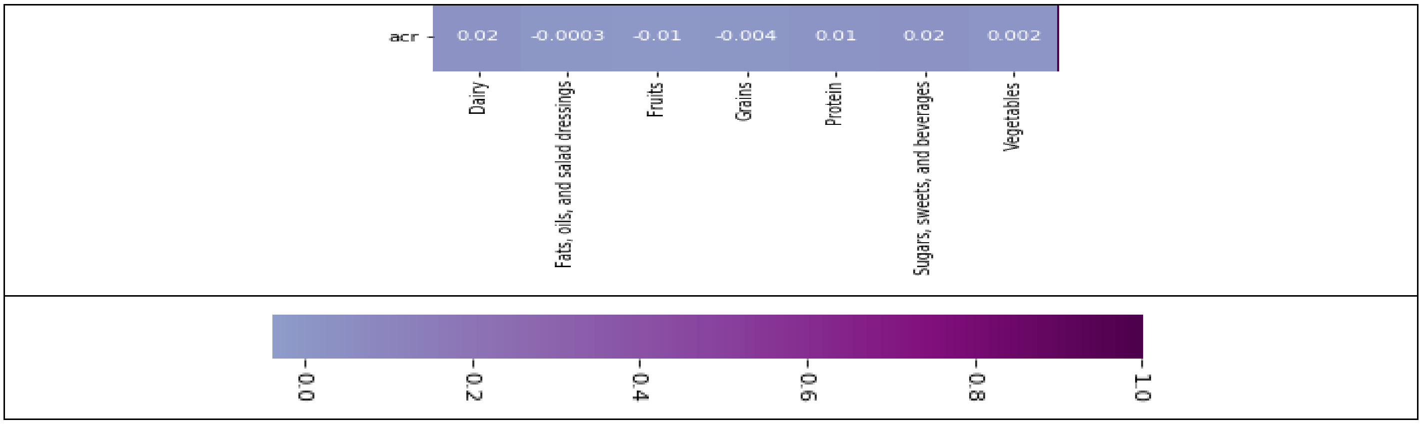
Effect on ACR using Regression Techniques: Effect on ACR was not found to be very significant in any of the experiments in this set. However, some food groups and/or nutrients have higher affect than the others

Regression: Food Groups and Nutrients Combined

Combined

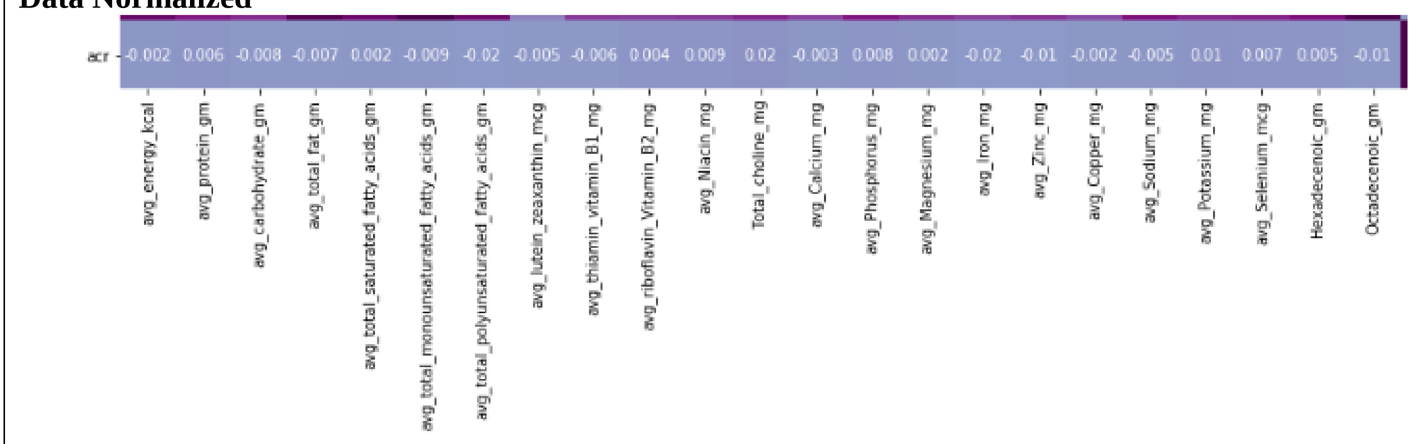


Regression: Food Groups:

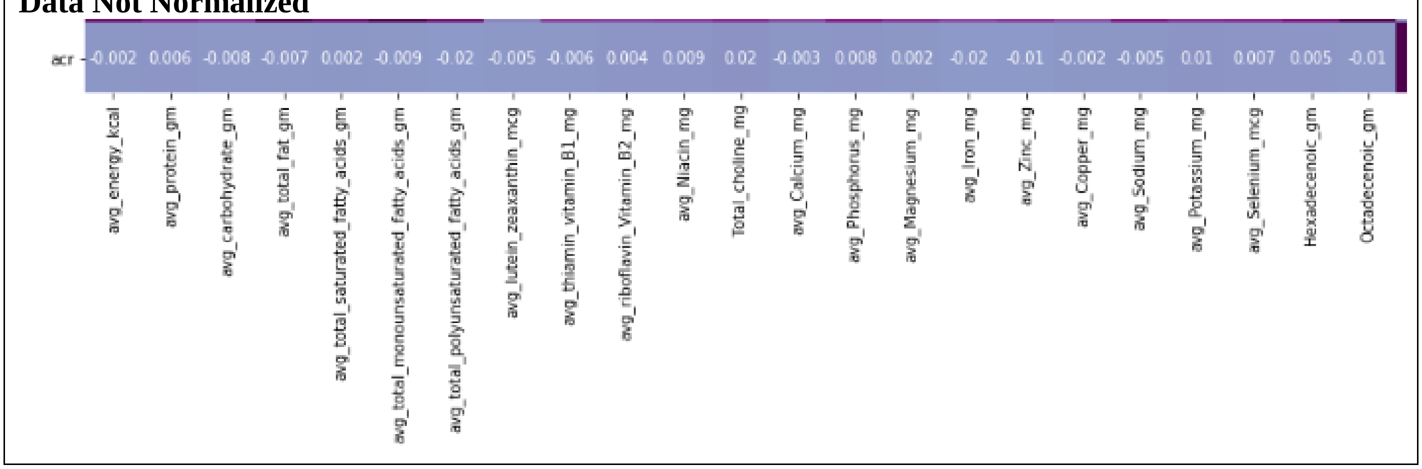


Regression: Nutrients and ACR:

Data Normalized

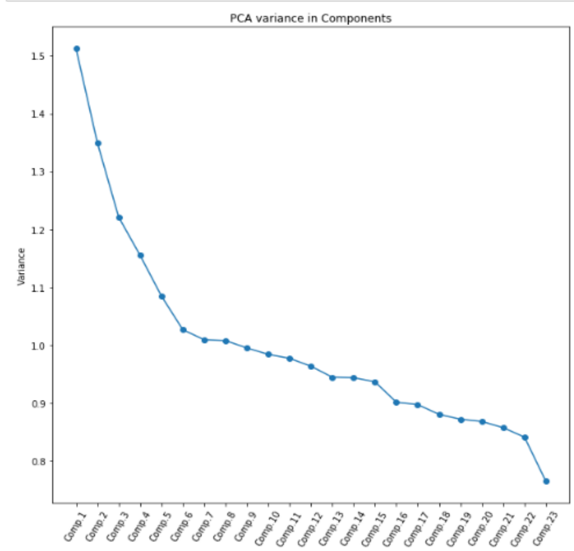


Data Not Normalized



Set 5 Food Subgroups and ACR Values

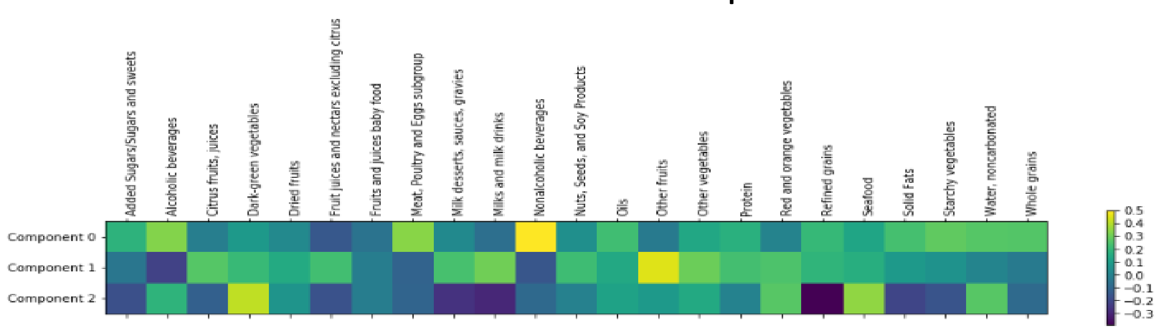
Experiment: Use PCA to identify food subgroups contributing in the dataset



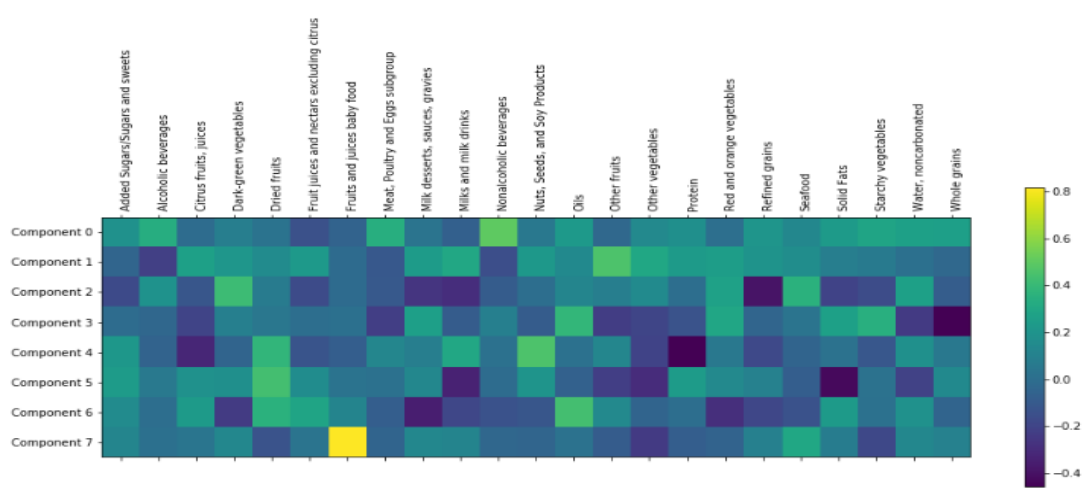
Importance of components:

	sdev	varprop	cumprop
PC1	1.229856	0.065763	0.065763
PC2	1.161816	0.058688	0.124451
PC3	1.105082	0.053096	0.177547
PC4	1.074889	0.050234	0.227781
PC5	1.041649	0.047175	0.274956
PC6	1.013388	0.044650	0.319606
PC7	1.004753	0.043893	0.363499
PC8	1.004013	0.043828	0.407327
PC9	0.997678	0.043277	0.450603
PC10	0.992131	0.042797	0.493400
PC11	0.988639	0.042496	0.535896
PC12	0.981714	0.041903	0.577799
PC13	0.972047	0.041082	0.618880
PC14	0.971623	0.041046	0.659926
PC15	0.967917	0.040733	0.700659

Contributions to Three Components

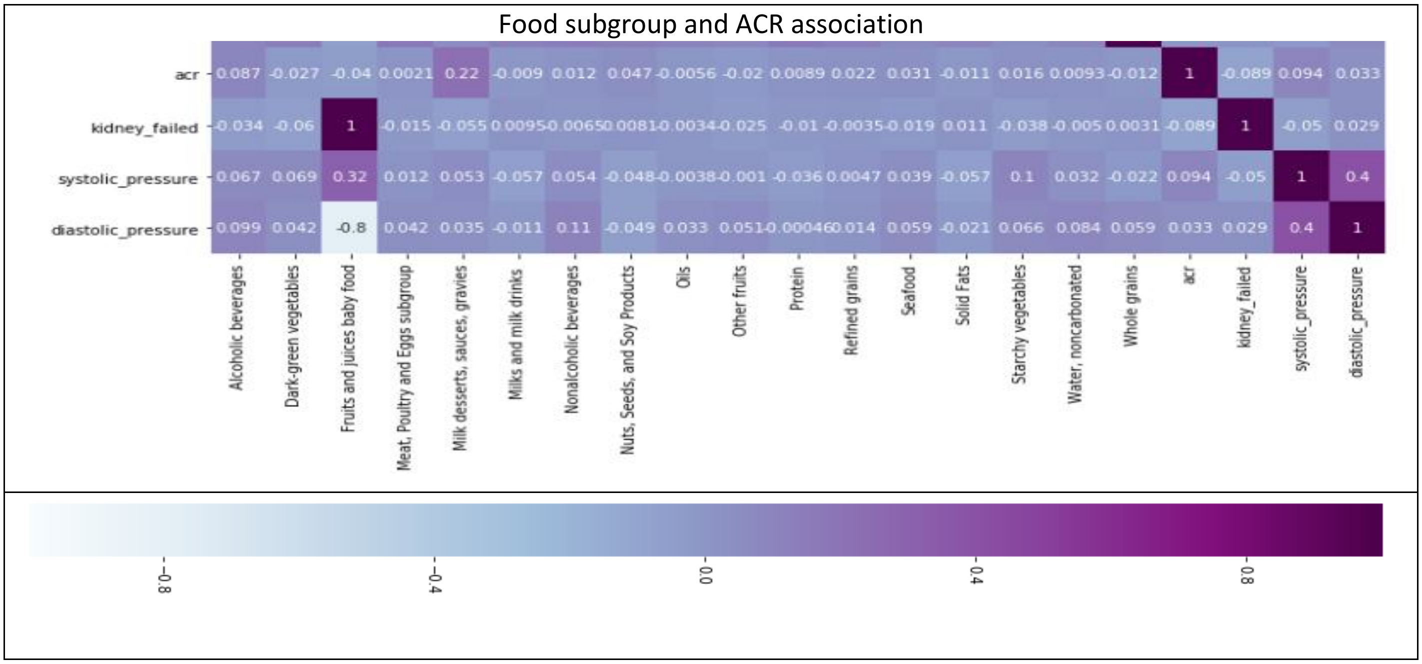


Contributions to 7 Components:



Experiment: Regression and Correlation:

'Milk desserts, Sauces, Gravies', and Alcoholic beverages show correlation with ACR more than the other food subgroups. However, the correlation can be seen to be negligible.



Set 6 Experiment using Machine Learning Approaches: ACR Values and Food Subgroups

Goal: If the ACR value can be predicted using the food intake patterns.

Applied Machine Learning (ML) Approaches such as Regression, Polynomial Regression, Random Forest Regression, Bayesian, and 10 fold cross validation on Food Subgroups dataset. Only the food subgroups that were found to be important using PCA were used for the ML approaches.

Target Variables:

Absolute ACR values and ACR Category were used as the target variables. For ACR category, < 30 is assigned to class 0, and > 30 is assigned to class 1.

Outcome:

The best test set accuracies are found using the approaches such as: 10 Fold Cross Validation Polynomial Regression, Polynomial Bayesian with Cross Validation (68%), Polynomial Regression (57%), Bayesian on Polynomial fit (41%), Cross Validation with Polynomial Random Forest Regression (21%)

A list of the best performing approaches and the outcome are provided below. Complete list can be seen on outcome_machine_learning_for_acr_values.xlsx

Feature Data	Target	Approach used	MSE Train	MSE Test	RMSE Train	RMSE Test	R2 Score Train	Accuracy: Test R2 Score if not mentioned
Data Not Normalized	ACR Value	10 Fold Cross Validation Polynomial Regression						-0.957 - cross_val_score
Data Not Normalized	ACR Value	Polynomial Bayesian with Cross Validation						-0.682 - cross_val_score
Data Not Normalized	ACR Value	Polynomial Regression	90965	52946	301	301	0.359	-0.579 – r2_score – test data
Data Not Normalized	ACR Value	Bayesian on Polynomial fit	93047	47431	305	305	0.344	-0.414 r2_score on test data

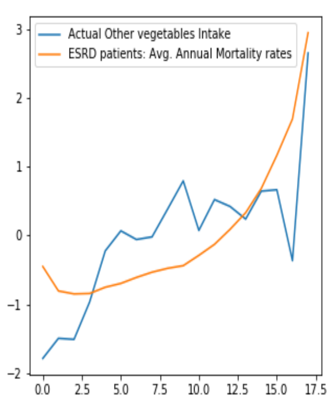
Outcome when ACR category is used:
After regression, y > 0.5 is assigned to category 1. Test accuracies are 88%

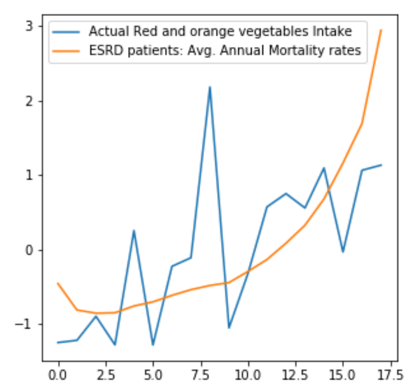
Data	Target	Appoach	Train Confusion (total, correct : [TP, FN, FP, TN])	Test Confusion
Not Normalized	acr_category	LinearRegression	6927,87 :[6032,0,895,0]	770, 88: [692, 0, 78, 0]
Normalized	acr_category	LinearRegression	6927, 87:[6032,0,895,0]	770, 88: [692, 0, 78, 0]

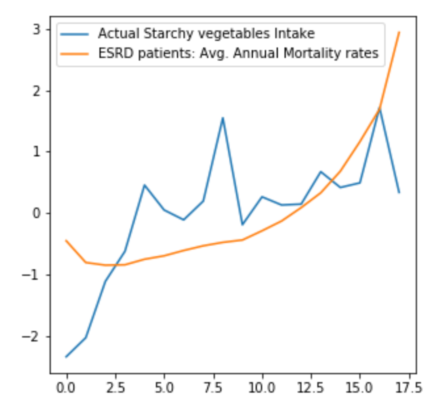
Appendix

Line chart plots for set 2: Experiment: Use Regression to find correlations to Mortality rates with food subgroups.

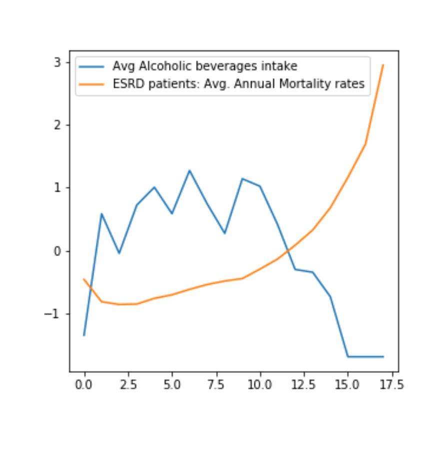
Positively:

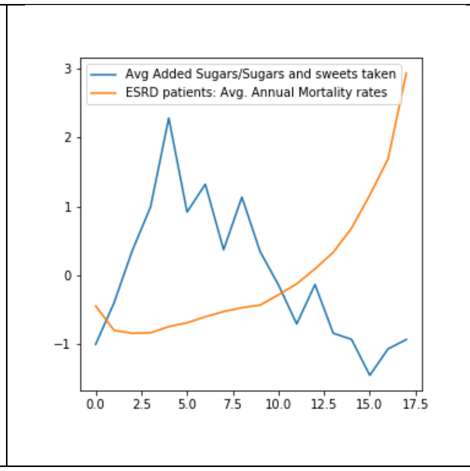


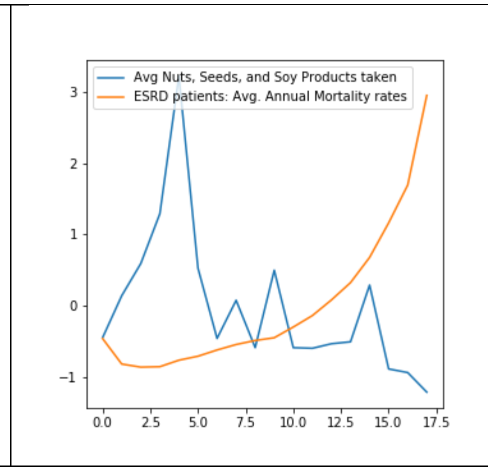




Negatively:







On Github:

Repository:

public-data-code-the-effect-of-dietary-patterns-on-the-mortality-and-survival-of-ckd-patients

https://github.com/sayed-ahmed-canada/public-data-code-the-effect-of-dietary-patterns-on-the-mortality-and-survival-of-ckd-patients

Python Files: Shows the plots and sample data:

Python files (.ipynb and .py) that will also show the plots and outputs.

https://github.com/sayed-ahmed-canada/public-data-code-the-effect-of-dietary-patterns-on-the-mortality-and-survival-of-ckd-patients/tree/master/Python-or-R-codes

Submitted on D2l: submitting-july-17th-experiment-phase-code-and-data-primarily.zip will have Python codes and data (major ones) as well as SQL Stored procedures used for data processing. Github repository will have more (all the important one) data.

Submitted zip files will also have Python files in the same folder

Analysis using Excel and XLStat (This is an older version):

Food subgroups and ACR. Shows similar output to Python files. The information can be representative though will not reflect the latest status/output/result
https://github.com/sayed-ahmed-canada/public-data-code-the-effect-of-dietary-patterns-on-the-mortality-and-survival-of-ckd-patients/blob/master/phase-experiment-excel-xlstat-analysis/food-groups/upto-date-jul-06-applying_pca_and_regression_on_this_merged_food_items_and_nutrients_file.xlsm

References

1. What Is Chronic Kidney Disease?
2. Jaimon T. Kelly, Suetonia C. Palmer, Shu Ning Wai, Marinella Ruospo, Juan-Jesus Carrero, Katrina L. Campbell, and Giovanni F. M. Strippol Healthy Dietary Patterns and Risk of Mortality and ESRD in CKD: A Meta-Analysis of Cohort Studies
3. Chen X, Wei G, Jalili T, Metos J, Giri A, Cho ME, Boucher R, Greene T, Beddhu S: The associations of plant protein intake with all-cause mortality in CKD. Am J Kidney Dis 67: 423–430, 2016 (26)

4. Gutie´rrez OM, Muntner P, Rizk DV, McClellan WM, Warnock DG, Newby PK, Judd SE: Dietary patterns and risk of death and progression to ESRD in individuals with CKD: A cohort study. Am J Kidney Dis 64: 204–213, 2014 (27)
5. Huang X, Jime´nez-Moleo´n JJ, Lindholm B, Cederholm T, Arnold˙v J, Rise´rus U, Sjo¨gren P, Carrero JJ: Mediterranean diet, kidney function, and mortality in men with CKD. Clin J Am Soc Nephrol 8: 1548–1555, 2013 (28)
6. Muntner P, Judd SE, Gao L, Gutie´rrez OM, Rizk DV, McClellanW, Cushman M, Warnock DG: Cardiovascular risk factors in CKD associated with both ESRD and mortality. J Am Soc Nephrol 24: 1159–1165, 2013 (29)
7. Ricardo AC, Madero M, Yang W, Anderson C, Menezes M, Fischer MJ, Turyk M, Daviglus ML, Lash JP: Adherence to a healthy lifestyle and all-cause mortality in CKD. Clin J Am Soc Nephrol 8: 602–609, 2013 (30)
8. Tsuruya K, Fukuma S, Wakita T, Ninomiya T, Nagata M, Yoshida H, Fujimi S, Kiyohara Y, Kitazono T, Uchida K, Shirota T, Akizawa T, Akiba T, Saito A, Fukuhara S: Dietary patterns and clinical outcomes in hemodialysis patients in Japan: A cohort study. PLoS One 10: e0116677, 2015 (31)
9. Ricardo AC, Anderson CA, Yang W, Zhang X, Fischer MJ, Dember LM, Fink JC, Frydrych A, Jensvold NG, Lustigova E, Nessel LC, Porter AC, Rahman M, Wright Nunes JA, Daviglus ML, Lash JP; CRIC Study Investigators: Healthy lifestyle and risk of kidney disease progression, atherosclerotic events, and death in CKD: Findings from the Chronic Renal Insufficiency Cohort (CRIC) Study. Am J Kidney Dis 65: 412–424, 2015 (17)
10. National Health and Nutrition Examination Survey
11. https://health.gov/dietaryguidelines/2015/guidelines/chapter-2/a-closer-look-at-current-intakes-and-recommended-shifts/
12. Food Code Numbers and the Food Coding Scheme
13. VEGETABLE SUBGROUPS
14. Key Concepts About the USDA Food Coding Scheme:
15. Appendix 3. USDA Food Patterns: Healthy U.S.-Style Eating Pattern.
16. United states Renal Data System (USRDS). 2018 ADR Reference Tables:
17. 2018 ADR Chapters.
18. Documentation and Dataset:
19. 2018 USRDS Annual Data Report: Executive Summary
20. DIETARY GUIDELINES FOR AMERICANS 2015-2020
21. https://health.gov/dietaryguidelines/dga95/9DIETGUI.HTM
22. Mortality and Causes of Death
23. Tong A, Chando S, Crowe S, Manns B, Winkelmayer WC, Hemmelgarn B, Craig JC: Research priority setting in kidney disease: A systematic review. Am J Kidney Dis 65: 674–683, 2015
24. Lin J, Fung TT, Hu FB, Curhan GC: Association of dietary patterns with albuminuria and kidney function decline in older white women: A subgroup analysis from the Nurses' Health Study. Am J Kidney Dis 57: 245–254, 2011
25. Taylor EN, Fung TT, Curhan GC: DASH-style diet associates with reduced risk for kidney stones. J Am Soc Nephrol 20: 2253–2259, 2009
26. Liu, Hao-Wen; Tsai, Wen-Hsin; Liu, Jia-Sin; Kuo, Ko-Lin. 2019. "Association of Vegetarian Diet with Chronic Kidney Disease." *Nutrients* 11, no. 2: 279.
27. Golaleh Asghari, Mehrnaz Momenan, Emad Yuzbashian, Parvin MirmiranEmail author and Fereidoun Azizi. Dietary pattern and incidence of chronic kidney disease among adults: a population-based study
28. Tanushree Banerjee1 , Deidra C. Crews2 , Delphine S. Tuot3 , Meda E. Pavkov4 , Nilka Rios Burrows4 , Austin G. Stack5 , Rajiv Saran6,7 , Jennifer Bragg-Gresham6 and Neil R. Powe1,8 ; for the Centers for Disease Control and Prevention Chronic Kidney Disease Surveillance Team9 Poor accordance to a DASH dietary pattern is associated with higher risk of ESRD among adults with moderate chronic kidney disease and hypertension
29. Jacek R., Beata F., Aleksandra C., Anna G.The Effect of Diet on the Survival of Patients with Chronic Kidney Disease. *Nutrients* 2017, *9*(5), 495; https://doi.org/10.3390/nu9050495
30. https://www.kidney.org/news/one-seven-american-adults-estimated-to-have-chronic-kidney-disease

31. https://www.medicalnewstoday.com/articles/282929.php
32. Five Stages of CKD. https://www.davita.com/education/kidney-disease/stages
33. Goal: How to Identify the Most Important Predictor Variables in Regression Models https://blog.minitab.com/blog/adventures-in-statistics-2/how-to-identify-the-most-important-predictor-variables-in-regression-models
34. How to Interpret Regression Analysis Results: P-values and Coefficients https://blog.minitab.com/blog/adventures-in-statistics-2/how-to-interpret-regression-analysis-results-p-values-and-coefficients
35. Regression Analysis: How to Interpret the Constant (Y Intercept) https://blog.minitab.com/blog/adventures-in-statistics-2/regression-analysis-how-to-interpret-the-constant-y-intercept
36. How to Compare Regression Slopes: How to statistically test the difference between regression slopes and constants https://blog.minitab.com/blog/adventures-in-statistics-2/how-to-compare-regression-lines-between-different-models
37. How Do I Interpret R-squared and Assess the Goodness-of-Fit? https://blog.minitab.com/blog/adventures-in-statistics-2/regression-analysis-how-do-i-interpret-r-squared-and-assess-the-goodness-of-fit
38. How High Should R-squared Be in Regression Analysis? https://blog.minitab.com/blog/adventures-in-statistics-2/how-high-should-r-squared-be-in-regression-analysis
39. How to Interpret a Regression Model with Low R-squared and Low P values https://blog.minitab.com/blog/adventures-in-statistics-2/how-to-interpret-a-regression-model-with-low-r-squared-and-low-p-values
40. Use Adjusted R-Squared and Predicted R-Squared to Include the Correct Number of Variables https://blog.minitab.com/blog/adventures-in-statistics-2/multiple-regession-analysis-use-adjusted-r-squared-and-predicted-r-squared-to-include-the-correct-number-of-variables
41. How to Interpret S, the Standard Error of the Regression https://blog.minitab.com/blog/adventures-in-statistics-2/regression-analysis-how-to-interpret-s-the-standard-error-of-the-regression
42. What Is the F-test of Overall Significance in Regression Analysis? https://blog.minitab.com/blog/adventures-in-statistics-2/what-is-the-f-test-of-overall-significance-in-regression-analysis
43. Understanding Analysis of Variance (ANOVA) and the F-test https://blog.minitab.com/blog/adventures-in-statistics-2/understanding-analysis-of-variance-anova-and-the-f-test
44. How to Compare Regression Slopes https://blog.minitab.com/blog/adventures-in-statistics-2/how-to-compare-regression-lines-between-different-models
45. How to Present and Use the Results to Avoid Costly Mistakes, part 1 https://blog.minitab.com/blog/adventures-in-statistics-2/applied-regression-analysis-how-to-present-and-use-the-results-to-avoid-costly-mistakes-part-1
46. How to Identify the Most Important Predictor Variables in Regression Models https://blog.minitab.com/blog/adventures-in-statistics-2/how-to-identify-the-most-important-predictor-variables-in-regression-models
47. How to Interpret your Regression Results http://sitestree.com/how-to-interpret-your-regression-results/