

Results

Effect of dietary patterns on Chronic Kidney Disease (CKD) Measures (ACR), and
on the mortality and survival of 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

Contents

Results.....	3
Food Groups and Mortality.....	3
Food Subgroups and Mortality	4
Mortality Study with non-aggregated Data	4
Food Groups, Food Nutrients, and Albumin Creatinine Ratio (ACR) Association.....	5
Food Subgroups and Albumin Creatinine Ratio (ACR) association	5
Predictability of ACR values based on Dietary patterns	5
Outcome when ACR Values are used as the Target.....	6
Outcome when ACR Category is Used as the Target	6
Appendix	7
More plots to show positive or negative correlations:.....	7
ML Approach Results	10
Code and Data.....	10
References	11

Results

Food Groups and Mortality

Experiments with aggregated (both NHANES and USRDS data) data to find associations between food groups and CKD mortality using Principal Component Analysis (PCA) and Regression show that Grains (-0.84) and Fruits (-0.43) have negative correlations with CKD mortality i.e. mortality is high for the patients who took significantly lower amount of Grains and Fruits than recommended amounts. Data exploration (plots below) also reflects the negative relation. As the correlation for fruits is -0.43 i.e. not very high, hence, Fruits can be thought of moderately associated.

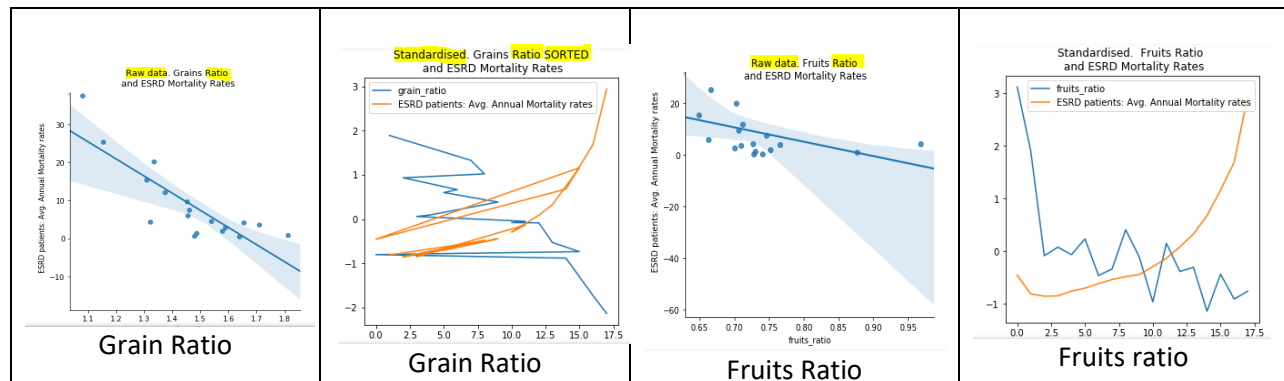


Fig: Food Groups and Mortality – Negative Relations

Vegetables show positive (0.58) correlation i.e. mortality is high for the patients who took more vegetables. As the correlation is 0.58, hence, this is not a very strong conclusion. Data shows this correlation in older adults. Even though ratios of food intake amount to high end of recommended amounts (actual intake amounts also show positive relation) were used; age might have biased the correlation. This does not show conformity with the general recommendation to take more vegetables for CKD patients. However, as experiments with food subgroups show that vegetable subgroups such as Other vegetables (0.68), Red and Orange vegetables (0.55), and Starchy vegetables (0.44) show positive correlations that have an impact on the Vegetable group correlation. Though more vegetable intake is a general recommendation for CKD patients; however, certain vegetables with more carbohydrates (and sugars) such as Starchy Vegetables as well as Vegetables with more Potassium and Calcium such as Tomatoes, and Spinach are recommended to take less. Considering the subgroups, the moderate positive correlation as this study found for vegetables food group is in conformity with the general recommendations for CKD patients.

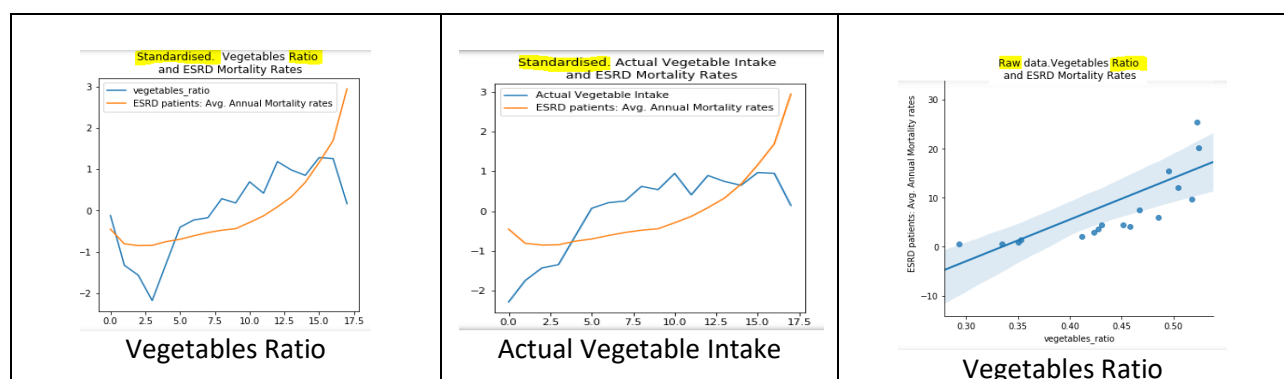
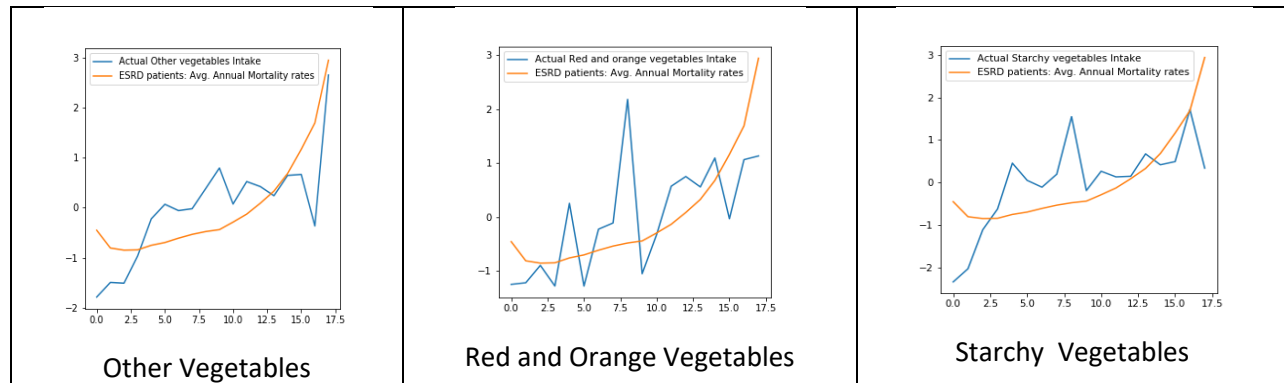


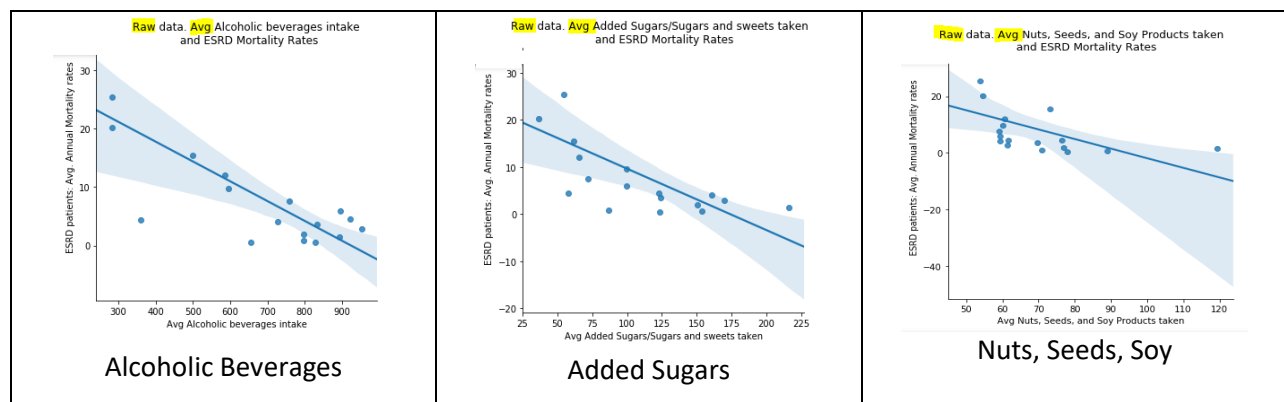
Fig: Food Groups and Mortality – Positive Relations

Food Subgroups and Mortality

Experiments with aggregated (both NHANES and USRDS data) data to find associations between food subgroups and CKD mortality using PCA and Regression show that Other vegetables (0.68), Red and orange vegetables (0.55), and Starchy vegetables (0.44) have positive correlations with mortality i.e. mortality is low when the intake amounts are low, and mortality is high when intake amounts are high. Data Exploration plots as given below also show the positive relations.



Food subgroups such as Alcoholic beverages (-0.79), Added Sugars/Sugars and sweets (-0.64), Whole grains (-0.61), and 'Nuts, Seeds, and Soy Products' (-0.55) show the most negative correlations with CKD mortality. Data exploration also shows negative correlations as shown in the charts below. These outcomes can also be seen consistent with current knowledge except for Sugars. Prevalence of Stage 3 CKD is lower in Alcohol Drinkers than non-drinkers [2], Nuts being Phosphorous rich and Whole Grains being Potassium rich are detrimental to CKD patients and can cause higher mortality when taken more.



Mortality Study with non-aggregated Data

Experiments where mortality rates based on ages were brought for each NHANES survey row (i.e. not aggregated), the following food subgroups show more **positive** correlations than others: Fats, Eggs, Other vegetables, Nonalcoholic beverages, White potatoes and Puerto Rican starchy vegetables, Tomatoes and tomato mixtures, Oils, Deep-yellow vegetables respectively. In the study the following subgroups showed more **negative** correlations than others: Grain mixtures, frozen plate meals, soups, Crackers and salty snacks from grain products, Milks and milk drinks, Sandwiches with Meat, Poultry,

and fish, Poultry. Though the correlation numbers in this study are very low and negligible, however, the positive and negative correlations are consistent with current knowledge [1, 2, 3, 4, 5, 6, 7, 8, 9, 10].

Food Groups, Food Nutrients, and Albumin Creatinine Ratio (ACR) Association

The experiments using PCA and Regression showed negligible effect on ACR with food groups and subgroups. However, some food groups and/or nutrients such as Dairy, and 'Sugars, Sweets, and Beverages' have higher and positive though negligible (0.02) effect than the others where Fruits (-0.01) showed negative effect. For nutrients, Poly unsaturated fatty acids (-0.02), and Iron (-0.02) have negative correlation where Choline (0.02) showed better positive correlation than others. Findings for Choline matches with medical knowledge [1]

As the correlations are not significant further analysis can be done on the data especially for food groups and nutrients that are found important (using PCA) in the data as provided below:

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

Food Subgroups and Albumin Creatinine Ratio (ACR) association

The experiments showed 'Milk desserts, Sauces, Gravies' (0.22), and Alcoholic Beverages (0.087) have more positive correlations with ACR than the other food subgroups i.e. taking more of these food subgroups results higher ACR values. Research by Uehara et al. [8] also shows that excessive Alcohol consumption can cause Proteinuria/Albuminuria (high ACR). Nettleton et al. [9] found that high fat dairy can be linked to high ACR values where low-fat dairy is not strongly linked to high ACR values. However, the correlation as this research found is very low. Low values might still explain a correlation where ACR values might depend on other factors in together than only these food subgroups. Fruits and juicy baby foods show negative correlation (-0.04) though not significant i.e. taking high amount does not increase ACR values that matched with current knowledge [10].

Predictability of ACR values based on Dietary patterns

Experiments were conducted to discover if ACR values can be predicted using the food intake patterns. Machine Learning (ML) Approaches such as Regression, Polynomial Regression, Random Forest Regression, Bayesian prediction with or without 10 fold cross validations were applied on Food Subgroups intake 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, $ACR < 30$ is assigned to class 0, and $ACR > 30$ is assigned to class 1. ACR values less than 30 indicate no CKD, where ACR values between 30 and 300 indicate moderate CKD. ACR value greater than 300 is considered severe CKD.

Outcome when ACR Values are used as the Target

The best test set accuracies were found using the approaches such as: 10 Fold Cross Validations with Polynomial Regression (95%), Polynomial Bayesian with 10 fold Cross Validations (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 in the appendix.

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 as the Target

After regression, $y > 0.5$ is assigned to category 1 (CKD), others were assigned to category 0. Test accuracies are 88%. Format for confusion matrix data in the table below: (total, % Correct : [TP, FN, FP, TN])

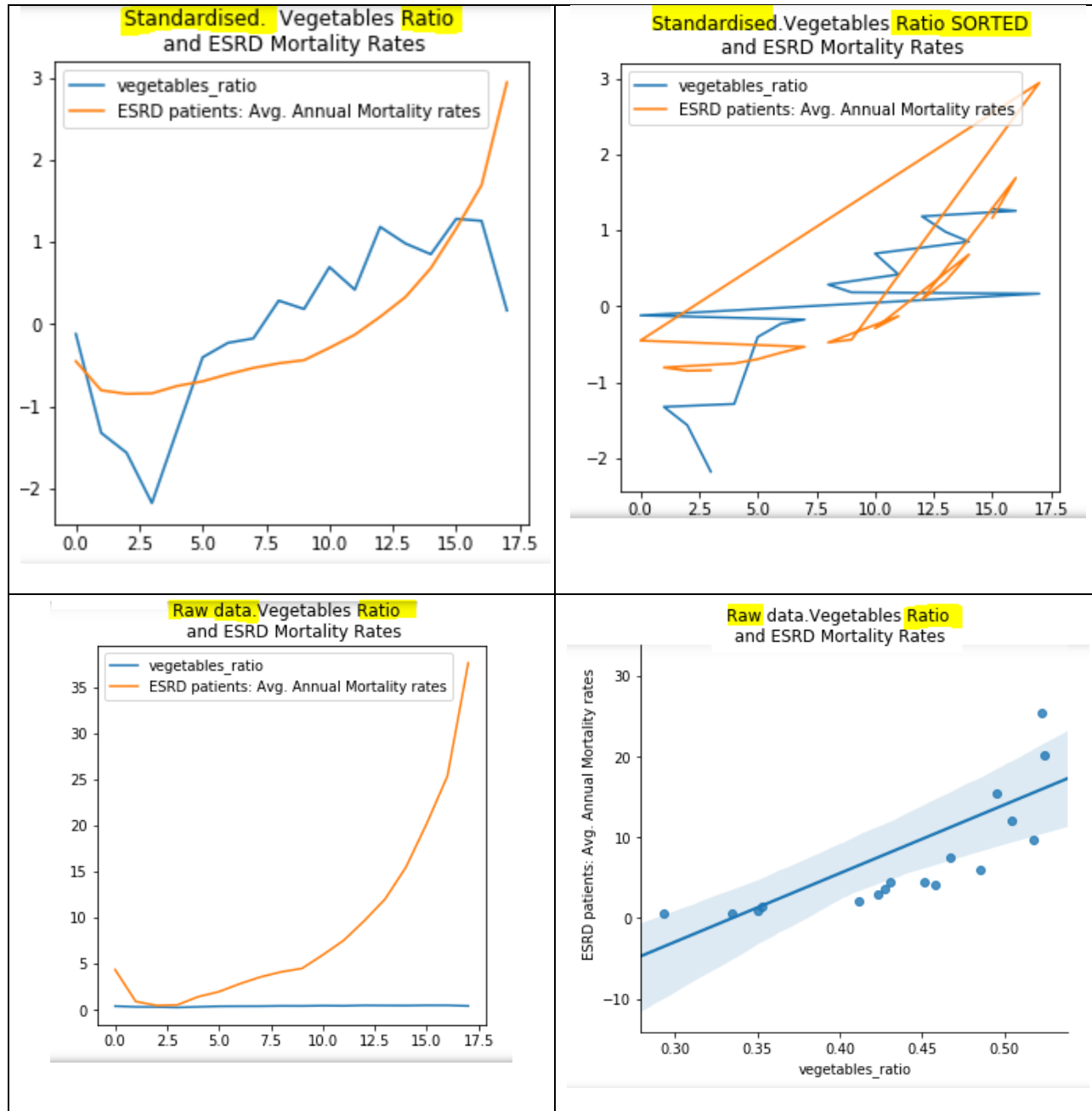
Data	Target	Approach	Train Confusion Matrix	Test Confusion Mat.
Not Normalized	acr_category	Linear Regression	6927, 87 :[6032,0,895,0]1	770, 88: [692, 0, 78, 0]
Normalized	acr_category	Linear Regression	6927, 87:[6032,0,895,0]	770, 88: [692, 0, 78, 0]

The high prediction accuracies might relate to the fact that only 10 to 14% population has ACR > 30. However, as cross validations also show high accuracy, it can be concluded that ACR values can be well predicted using Machine Learning approaches especially with 10 Fold Cross Validation Polynomial Regression having accuracy 95%

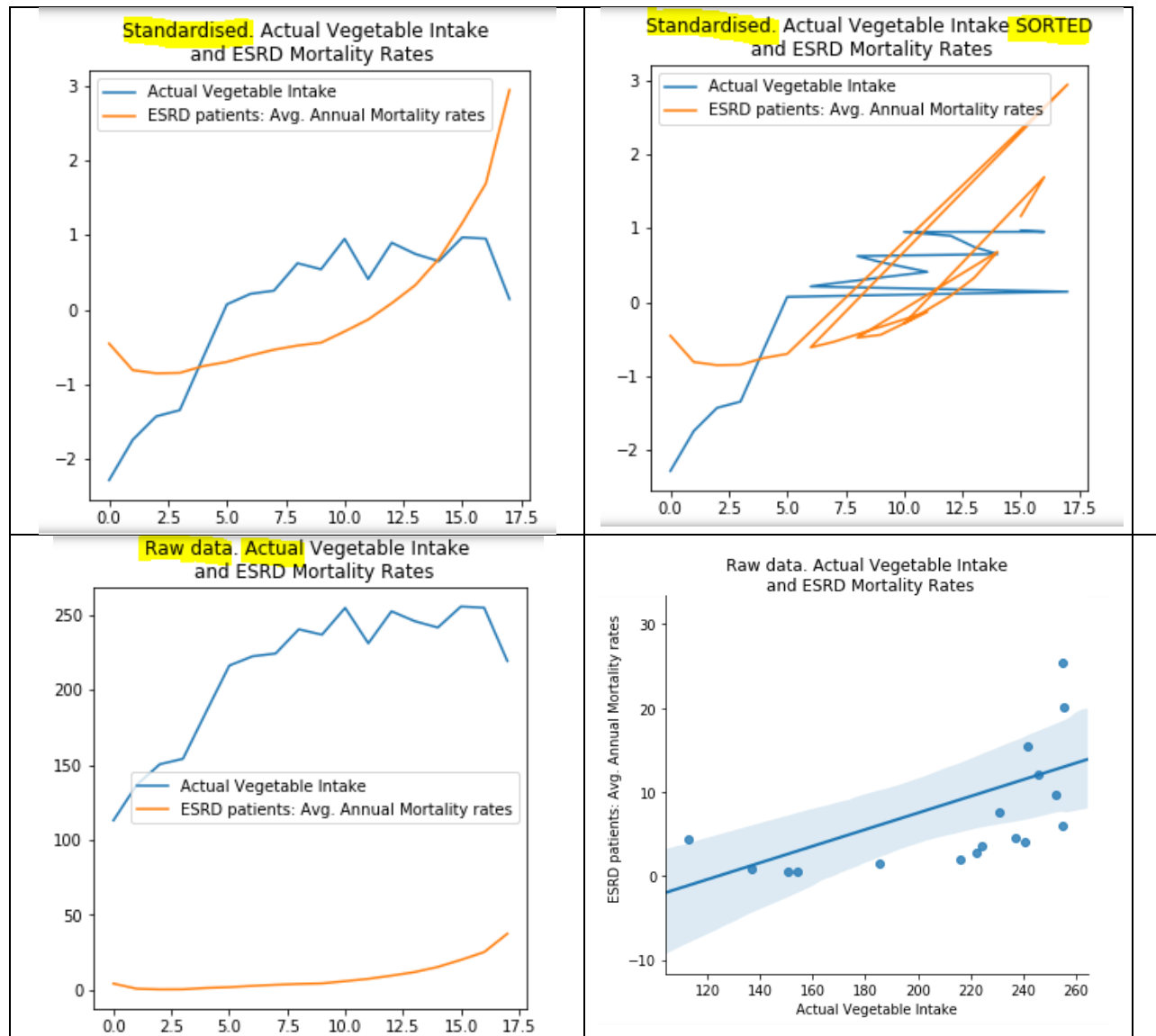
Appendix

More plots to show positive or negative correlations:

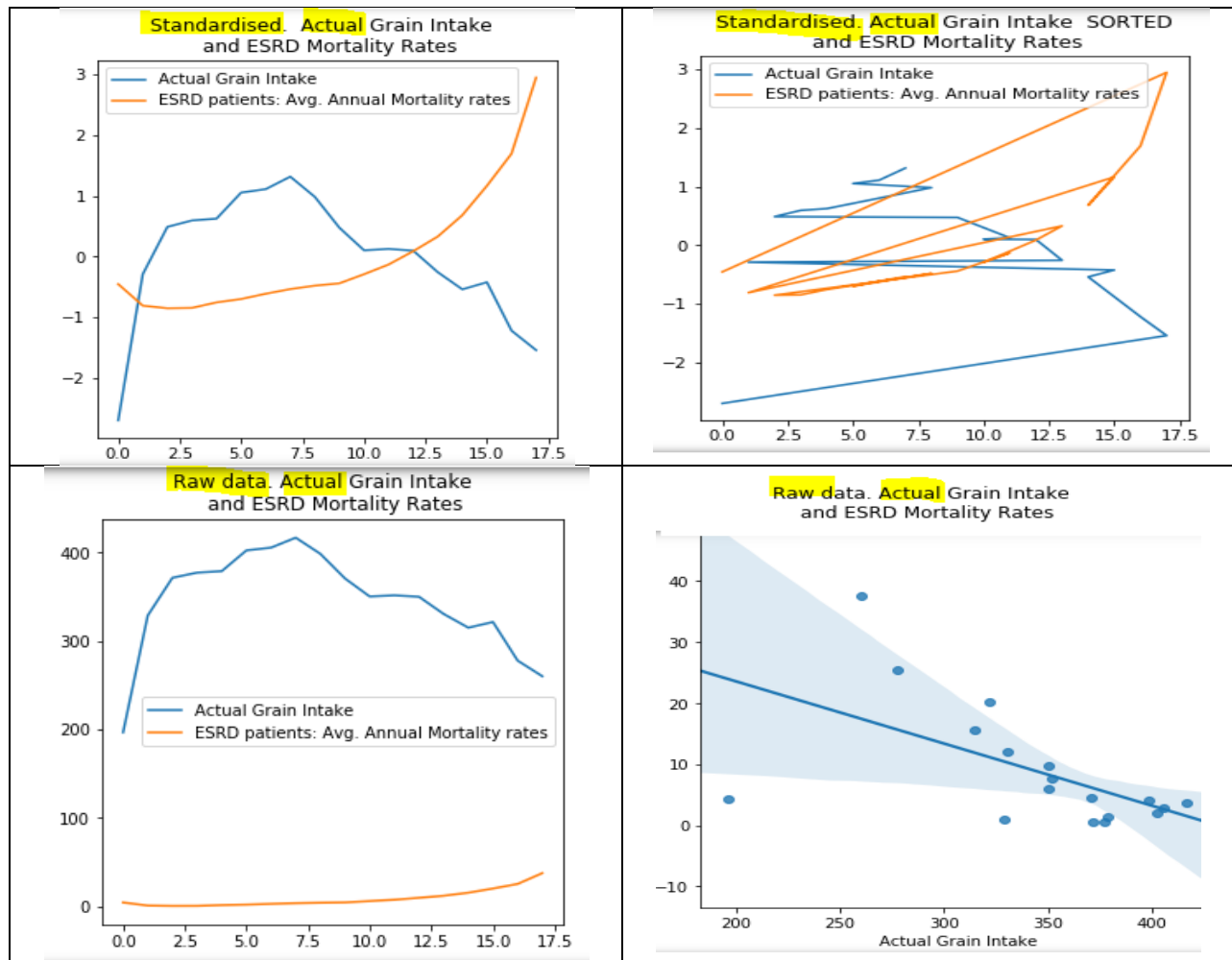
Vegetables and CKD Mortality



Food Subgroups and CKD Mortality



Grain and Mortality: Negative



Subgroup and Mortality



ML Approach Results

Data	Target	Approach used	MSE Train	MSE Test	RMSE Train	RMSE Test	R2 Score Train	Accuracy /performance: Test R2 Score
Not Normalized	ACR Value	10 Fold Cross Validation Polynomial Regression						-0.957
Not Normalized	ACR Value	Polynomial Bayesian with Cross Validation						-0.682
Not Normalized	ACR Value	Polynomial Regression	90965.000	52946.43	301.604	301.604	0.359	-0.579
Not Normalized	ACR Value	Bayesian on Polynomial fit	93047.912	47431.675	305.038	305.038	0.344	-0.414
Not Normalized	ACR Value	Cross Validation with Polynomial Random Forest Regressor						-0.394
Normalized	ACR Value	10 Fold Cross Validation with Linear Regression						-0.134
Not Normalized	ACR Value	10 Fold Cross Validation with Linear Regression						-0.134
Normalized	ACR Value	Linear Regression	139019.96	36128.194	372.854	190.074	0.020	-0.077
Not Normalized	ACR Value	Linear Regression	139019.96	36128.194	372.854	190.074	0.020	-0.077
Not Normalized	ACR Value	Bayesian	139091.64	35430.367	372.950	372.950	0.019	-0.056
Data Not Normalized	acr_category	10 Fold Cross Validation with Linear Regression						-0.053
Data Normalized	acr_category							-0.053
Data Not Normalized	ACR Value	Polynomial Random Forest Regression	77719.25	34344.340	278.780	278.780	0.452	-0.024
Data Not Normalized	acr_category	LinearRegression	0.112	0.092	0.335	0.303	0.005	-0.008
Data Normalized	acr_category	LinearRegression	0.112	0.092	0.335	0.303	0.005	-0.003

Code and Data

Code used for the analysis and the results can be found in the folder: **Python-or-R-codes**. Data as used by the codes will be under a subfolder of **Python-or-R-codes** named **data-for-code**. Other relevant data if any can be found under the folders: **data-for-analysis**, **data-to-help-with-analysis** (or on Github or in the previous submissions for exploration and experiment phases)

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

References

As used in earlier reports such as reports for exploration phase and reports for experiment phase. As well as all data and code used and submitted before.

1. Raul Fernandez-Prado, Raquel Esteras, Maria Vanessa Perez-Gomez, Carolina Gracia-Iguacel, Emilio Gonzalez-Parra, Ana B. Sanz, Alberto Ortiz, Maria Dolores Sanchez-Niño. Nutrients Turned into Toxins: Microbiota Modulation of Nutrient Properties in Chronic Kidney Disease. *Nutrients*. 2017 May; 9(5): 489. Published online 2017 May 12. doi: 10.3390/nu9050489 PMID: PMC5452219 <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5452219/>
2. Hsu, Y. H., Pai, H. C., Chang, Y. M., Liu, W. H., & Hsu, C. C. (2013). Alcohol consumption is inversely associated with stage 3 chronic kidney disease in middle-aged Taiwanese men. *BMC nephrology*, 14, 254. doi:10.1186/1471-2369-14-254
3. Rippe, J. M., & Angelopoulos, T. J. (2016). Relationship between Added Sugars Consumption and Chronic Disease Risk Factors: Current Understanding. *Nutrients*, 8(11), 697. doi:10.3390/nu8110697
4. Karalius VP, Shoham DA. Dietary sugar and artificial sweetener intake and chronic kidney disease: a review. *Adv Chronic Kidney Dis*. 2013; 20:157–164. doi: 10.1053/j.ackd.2012.12.005
5. Jaimon T. Kelly, Suetonia C. Palmer, Shu Ning Wai, Marinella Ruospo, Juan-Jesus Carrero, Katrina L. Campbell and Giovanni F. M. Strippoli Healthy Dietary Patterns and Risk of Mortality and ESRD in CKD: A Meta-Analysis of Cohort Studies. *CJASN* February 2017, 12 (2) 272-279; DOI: <https://doi.org/10.2215/CJN.06190616>
6. Jilian k. The 20 Best Foods for People With Kidney Problems. <https://www.healthline.com/nutrition/best-foods-for-kidneys>
7. Drinking Alcohol Affects Your Kidneys <https://www.kidney.org/news/kidneyCare/winter10/AlcoholAffects>
8. Uehara, S., Hayashi, T., Kogawa Sato, K., Kinuhata, S., Shibata, M., Oue, K., ... Hashimoto, K. (2016). Relationship Between Alcohol Drinking Pattern and Risk of Proteinuria: The Kansai Healthcare Study. *Journal of epidemiology*, 26(9), 464–470. doi:10.2188/jea.JE20150158
9. Nettleton, J. A., Steffen, L. M., Palmas, W., Burke, G. L., & Jacobs, D. R., Jr (2008). Associations between micro albuminuria and animal foods, plant foods, and dietary patterns in the Multiethnic Study of Atherosclerosis. *The American journal of clinical nutrition*, 87(6), 1825–1836. doi:10.1093/ajcn/87.6.1825
10. Jacobs, D. R., Jr, Gross, M. D., Steffen, L., Steffes, M. W., Yu, X., Svetkey, L. P., ... Sacks, F. (2009). The effects of dietary patterns on urinary albumin excretion: results of the Dietary Approaches to Stop Hypertension (DASH) Trial. *American journal of kidney diseases : the official journal of the National Kidney Foundation*, 53(4), 638–646. doi:10.1053/j.ajkd.2008.10.048. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2676223/>