# MGT 6203 Group Project Proposal

# **TEAM INFORMATION (1 point)**

### Team #: 68

# **Team Members:**

### 1. Team Member 1

a. Name: Gabriel Mink

b. GTID: 903738167: gmink3@gatech.edu

c. Personal Background: I've been a Data Scientist/Machine Learning Engineer for the last 4 years mostly in the biotech industry, I graduated from UCSD majoring in Biology and minoring in Computer Science, I've worked on purchase prediction recommender systems and multivariate financial forecasting models.

#### 2. Team Member 2

a. Name: Bella (Yifei) Ding

b. GTID: 903131776: yding302@gatech.edu

c. Personal Background: I recently became a Data Engineer in the tech industry. Previously I completed my degree in Mechanical Engineering and worked in the diesel automotive industry as a Test Engineer for 2 years. I have worked on some analytics related projects mostly in school. One previous research project was on building an image classification model for bacteria lab images.

#### 3. Team Member 3

a. Name: Vincent Pan

b. GTID: 903847411: vipan@gatech.edu

c. Personal Background: I've been a Data Scientist since graduating university in Actuarial Studies and Mathematics. Currently, I live in the San Francisco Bay Area and am working to understand how WhatsApp can better serve large businesses and deliver communications they have with their customers.

#### 4. Team Member 4

a. Name: Nikolos Lahanis

b. GTID: 903674177: nlahanis3@gatech.edu

c. Personal Background: A recent mechanical engineering undergrad from NC State University, I currently live and work in the San Francisco Bay Area as an Intelligent Automation Consultant. Previous analytics projects I have worked on include an Investment-Grade Asset Ranking System and a Real Estate Property Valuation Tool.

#### 5. Team Member 5

a. Name: Rahul Sati

b. GT Id: 903549883: rsati3@gatech.edu

c. Personal Background: I am currently working as a Data Science Manager with a leading e-commerce company. I have completed my previous Masters in Industrial Engineering, focusing on operations research and have done my Bachelors in Mechanical Engineering. I have done data science projects in the area of forecasting, experimentation (A/B Testing) and simulation (simulating network design for Walmart).

# **OBJECTIVE/PROBLEM (5 points)**

**Project Title: Predict Blood Pressure Through Various Factors** 

# **Background Information on chosen project topic:**

Nowadays, as the awareness of maintaining health arises, more and more people started to pay more attention to a healthier lifestyle by, for example, choosing a healthier diet. Therefore, our team would like to investigate how much influence different factors have on a person's blood pressure through a National Health and Nutrition Examination Survey (NHANES) dataset. The modeling result could be very helpful to be future reference when providing health suggestions for patients.

### Problem Statement (clear and concise statement explaining purpose of your analysis and investigation):

This project will investigate the effect various factors, such as nutritional intake, diet and demographics, have on a person's blood pressure, in order to find out which factor(s) will have significant influence. Then this project will also predict blood pressure through those chosen factors.

### **State your Primary Research Question (RQ):**

Does blood pressure (or other health conditions/diseases) in the adults and children in the United States correlate with nutritional intake, diet, demographics, medical history and lifestyle? Can we predict the risk of having a disease based on these factors?

### Add some possible Supporting Research Questions (2-4 RQs that support problem statement):

- 1. Relationship of blood pressure with diet, demographics and other such factors?
- 2. Relationship of diabetes with diet, demographics and lifestyle?

Business Justification: (Why is this problem interesting to solve from a business viewpoint? Try to quantify the financial, marketing or operational aspects and implications of this problem, as if you were running a company, non-profit organization, city or government that is encountering this problem.)

The National Health and Nutrition Examination Survey (NHANES) is a program of studies designed to assess the health and nutritional status of adults and children in the United States. NHANES is a major program of the National Center for Health Statistics (NCHS). NCHS is part of the Centers for Disease Control and Prevention (CDC) and has the responsibility for producing vital and health statistics for the Nation.

In this project, the health, and nutritional data from NHANES is used to analyze and predict key factors leading to critical diseases. For example, high blood pressure can lead to heart diseases. According to the CDC heart disease is the leading cause of death for men, women, and people of most racial and ethnic groups in the United States. About 697,000 people in the United States died from heart disease in 2020, that's 1 in every 5 deaths. Heart disease costs the United States about \$229 billion each year, this includes the cost of healthcare services, medicines, and lost productivity due to death.

Similarly, we can model other critical health factors and diseases using the given data. Such application of data science and machine learning in healthcare will improve the ability of public health organizations to predict the outbreak of disease, improve the prevention of disease and enhance the quality of life. It will also improve the accuracy and speed of identifying patients at highest risk of disease.

# **DATASET/PLAN FOR DATA (4 points)**

## Data Sources (links, attachments, etc.):

- National Health and Nutrition Examination Survey from Kaggle.com from the Centers of Disease Control and Prevention (CDC)
- [Online] https://www.kaggle.com/datasets/cdc/national-health-and-nutrition-examination-survey

## Data Description (describe each of your data sources, include screenshots of a few rows of data):

Title:	National Health and Nutrition Examination Survey
	Tradional Fredien and Tradition Examination Survey
Source:	<u>Kaggle</u>
Origin:	NHANES datasets from the Centers of Disease Control and Prevention (CDC)
Dataset date:	2013-2014
Number of Files/ Datasets	6
Description:	"The National Health and Nutrition Examination Survey (NHANES) is a program of studies designed to assess the health and nutritional status of adults and children in the United States. The survey is unique in that it combines interviews and physical examinations" From a quick first glance, each of the dataset refer to an individual by the column: seqn which we can use to join the datasets together
Primary Key Across Datasets	SEQN: Respondent Sequence Number

Dataset Ref	Dataset Name	Description and Screenshots (See Appendix for more rows)
1	demographic.csv [Data Dictionary]	Description: Demographic data of the individual. This includes Age (DMDHRAGE), Race (RIDRETH1), and Gender (RIAGENDR)

	Number of Columns: 47 Number of Rows: 10.2k	Example Rows:           SEGN SDDSRYPR RIDSTATR RIAGENDR RIDAGEYR RIDAGEYN RIDAGEYN RIDEXAGM DMCMILIZ DMCADFC DMDBORN4 DMCCITZN DMCYRSUS DMDEDUC           73557         8         2         1         69         4         4         1         2         1         1         1         1         1         1         1         1         1         1         3         3         3         1         11         1         1         1         1         1         3         3         3         1         11         1         1         3         3         3         1         11         1 <td< th=""></td<>
2	diet.csv [Data Dictionary] Number of Columns: 168 Number of Rows: 9.8k	Data on the individual's first day diet. This is a questionnaire on how much e.g. salt (DBD100), amount of food was eaten (DR1_300), and tap water was drunk (DR1_320Z). Example Rows:
3	labs.csv [Data Dictionary] Number of Columns: 224 Number of Rows: 9.8k	Description: Data on the Labs. Not all tests were conducted for all individuals. However, some examples of test/ detection include: Monocyte number (LBDMONO), Platelet count (LBXPLTSI), and Potassium (LBXSKSI)
4	questionnaire.csv [Data Dictionary] Number of Columns: 953 Number of Rows: 10.2k	Description: A questionnaire asking about the health and behaviors of the individual. This includes time spent outdoors (DED125), whether they have been diagnosed with osteoporosis (OSQ060), and whether they are taking prescribed medicine (BPQ050A).  Example Rows:    Sequare   AcD011A   ACD011B   ACD011B   ACD011C   ACD010   ACD110   ALQ101   ALQ120Q   ALQ120Q   ALQ130   ALQ141Q   ALQ141   ALQ141   ALQ151   ALQ160   BPQ020   BPQ030   I   T   T   T   T   T   T   T   T   T
5	examinations.csv [Data Dictionary] Number of Columns: 424 Number of Rows: 9.8k	Description: Data on the health examination for the individual. This includes questions on whether they are currently pregnant or breastfeeding (CSQQ241), grip test (MGDEXSTS), and Total abdominal fat mass (DXXTATM)  Example Rows:

			SCTM: PEASCCT1 BPXCHR BPAARM BP	ACSZ BPXPLS	BPXPULS BPXPTY	BPXML1 BPXSY1	BPXDI1 E	PAEN1 BP	XSY2 BPXI	DI2 BPAE	N2 BPXSY3	BPXDI3 BPAEN3
		SEQN PEASCST1 PEAS 73557 1	620 1	4 8		140 1	22 72	2	114	76	2 10	2 74 2
		73558 1	766 1	4 7			6 62	2	160	80	2 15	
		73559 1	665 1	4 6			10 90	2	140	76	2 14	6 80 2
		73560 1	803 1	2 64			08 38	2	102	34	2 10	
		73561 1 73562 1	949 1 1064 1	3 9: 5 60			6 86 60 84	2	134 158	88 82	1 14	
		73563 1	90 152		1	100		-	250	OL.		- 00 2
		73564 1	954 1	5 83			18 80	2	124	80	2 12	
		73566 1 73567 1	625 1 932 1	4 8t			28 74 10 78	2	124 142	72 78	2 11	
		73568 1	585 1	3 70			06 60	2	100	62	2 10	
		73570 1	710 1	2 78			02 44	2	104	42	2 10	
			1193 1 557 1	4 7: 3 80			24 68	2			2 12	
		73572 1 73573 1	557 1 599 1	3 80			88 54 84 62	2	88 94	54 60	2 9	
		73574 1	710 1	3 9			2 56	2	124	66	2 11	
		73575 1	291 110		1							
		73576 1 73577 1	584 1 750 1	4 6		130 1 140 1	18 68 18 74	2	110 118	62 74	2 11	
		73578 1	607 1	5 6	1 1		0 58	2	174	64	2 12	
6	medications.csv [Data Dictionary]	This inclu	<i>ion:</i> Data on a udes the reas	on th		•						taking.
6		This inclu		on th		•						taking.
6	[Data Dictionary]	This inclu	udes the reas nsomnia etc.)	on th		•						taking.
6	[Data Dictionary] Number of Columns: 13	This inclusions in the spasm, in	udes the reas nsomnia etc.) <i>Rows:</i>	on th	ey are p	orescri	bed	for (	e.g.	Mu	scle	
6	[Data Dictionary] Number of Columns: 13	This inclusions spasm, in Example	udes the reas nsomnia etc.) Rows:	on th	ey are p	orescri	bed	for (	e.g.	Mu	scle	
6	[Data Dictionary] Number of Columns: 13	This inclusions in the spasm, in Example	udes the reas nsomnia etc.) Rows: RODRUG RXDDRGICE 1 99999	on th	ey are p	orescri	bed	for (	e.g.	Mu	scle	
6	[Data Dictionary] Number of Columns: 13	This inclusions inclusions from the spasm, in Example SEQN RXDUS 73557 73557	nsomnia etc.) Rows:  Rows:  R RODRUG RXDDRGIER 1 99999 1 INSULIN d00262	on the	ey are p	orescri	bed	for (	e.g.	Mu RXDRSD	2 RXDRS	
6	[Data Dictionary] Number of Columns: 13	This incluspasm, in Example SEQN RXDUS 73557 73558	nsomnia etc.)  Rows:  ROBE RADDRUG RADDRGIER  1 99999 1 INSULIN d00262 1 GABAPEN d03182	on the	Ey are p XDDAYS RXDR 1460 E11 243 G25.8	orescri	bed	for (	e.g.	Mu RXDRSD petes mo	2 RXDRS	
6	[Data Dictionary] Number of Columns: 13	This incluspasm, in Example SEQN RXDUS 73557 73558 73558	ROWS: SE RXDDRUG RXDDRGIC F 1 99999 1 INSULIN d00262 1 GABAPEN d03182 1 INSULIN G d04538	on the	Ey are K XXDDAYS RXDR 1460 E11 243 G25.8 365 E11	orescri	bed	for (	e.g.  DRSD1 F  e 2 diab  tless lege 2 diab	Mu RXDRSD petes me gs syndro	2 RXDRS ellitus rome ellitus	D3 RXDCOUNT 2 2 2 4 4 4
6	[Data Dictionary] Number of Columns: 13	This incluspasm, in Example SEQN RXDUS 73557 73558	nsomnia etc.)  Rows:  ROBE RADDRUG RADDRGIER  1 99999 1 INSULIN d00262 1 GABAPEN d03182	on the	EY are p XXDDAYS RXDR 1460 E11 243 G25.8 365 E11 14 E11.2	orescri	bed	for (	PRSD1 F	Mu RXDRSD petes me gs syndro	2 RXDRS	D3 RXDCOUNT 2 2 2 4 4 4
6	[Data Dictionary] Number of Columns: 13	This incluspasm, in Example SEQN RXDUS 73557 73558 73558	ROWS: SE RXDDRUG RXDDRGIC F 1 99999 1 INSULIN d00262 1 GABAPEN d03182 1 INSULIN G d04538	on the	Ey are K XXDDAYS RXDR 1460 E11 243 G25.8 365 E11	orescri	bed	for (	ORSD1 F	Mu RXDRSD Detes mi gs syndro Detes mo	2 RXDRS ellitus rome ellitus	D3 RXDCOUNT 2 2 4 4 4 th k 4
6	[Data Dictionary] Number of Columns: 13	This incluspasm, in Example  SEQN RXDUS 73557 73557 73558 73558 73558	ROWS:  RE RXDDRUG RXDDRGICE 1 99999 1 INSULIN d00262 1 GABAPEN d03182 1 INSULIN G d04538 1 OLMESAR d04801	on the	EY are p XXDDAYS RXDR 1460 E11 243 G25.8 365 E11 14 E11.2	orescri	bed	for (	ORSD1 F	Mu RXDRSD petes me gs syndro petes me petes me rcholest	2 RXDRS ellitus rome ellitus ellitus wi erolemia	D3 RXDCOUN' 2 2 4 4 4 th k 4
6	[Data Dictionary] Number of Columns: 13	This inclusions are spasm, in Example  SEQN RXDUS 73557 73557 73558 73558 73558 73558 73558	ROWS:  SE RXDDRUG RXDDRGICE 1 99999 1 INSULIN d00262 1 GABAPEN d03182 1 INSULIN G d04538 1 OLMESAR d04801 1 SIMVASTA d00746	on the	EY are p XXDDAYS RXDR 1460 E11 243 G25.8 365 E11 14 E11.2 61 E78.0	orescri	bed	for (	PRSD1 Fine 2 diab	Mu RXDRSD petes mi gs syndr petes mi recholest petes mi	2 RXDRS ellitus rome ellitus ellitus wi erolemia ellitus	D3 RXDCOUN 2 2 4 4 th   4
6	[Data Dictionary] Number of Columns: 13	This incluspasm, in Example SEQN RXDUS 73557 73558 73558 73558 73558 73558 73558 73559	Modes the reases and somnia etc.)  Rows:  SE RXDDRUG RXDDRGILF 1 99999 1 INSULIN d00262 1 GABAPEN d03182 1 INSULIN G d04538 1 OLMESAR d04801 1 INSULIN A d04697 1 INSULIN G d04538	on the	EY ARE K  XDDAYS RXDR  1460 E11  243 G25.8  365 E11  4 E11.2  61 E78.0  365 E11	orescri	bed	Typ Res Typ Pur Typ Typ	PRSD1 Fine 2 diabitiless lege 2 diabitiless lege 2 diabitiles e 2 diabitiles 2 diab	Mu exxprsp petes mi gs syndi petes mi petes mi rcholest petes mi petes mi	2 RXDRS ellitus rome ellitus ellitus wi erolemia ellitus ellitus	D3 RXDCOUN 2 2 4 4 th   4 5 5
6	[Data Dictionary] Number of Columns: 13	This incluspasm, in Example SEQN RXDUS 73557 73558 73558 73558 73558 73559 73559 73559	ROWS:  SE RXDDRUG RXDDRGIC F  1 99999 1 INSULIN d00262 1 GABAPEN d03182 1 INSULIN G d04538 1 OLMESAR d04801 1 SIMVASTA d00746 1 INSULIN A d04697 1 INSULIN G d04538 1 PANCRELI d01002	ON the	EY ARE F  XDDAYS RXDR  1460 E11  243 G25.8  365 E11  14 E11.2  61 E78.0  365 E11  4380 E11  365 K86.9	orescri	bed	Typ Res Typ Pur Typ Typ Disc	PRSD1 F  DRSD1 F  ee 2 diab  titless leg  ee 2 diab  ee 2 diab  ee 4 diab  ee 4 diab  ee 5 diab  ease of	Mu exxDRSD petes mi gs syndi petes mi petes mi petes mi petes mi petes mi petes mi	2 RXDRS ellitus rome ellitus ellitus wi erolemia ellitus ellitus as, unspe	D3 RXDCOUNT 2 2 4 4 th l 4 5 5 cifit 5
6	[Data Dictionary] Number of Columns: 13	This incluspasm, in Example SEQN RXDUS 73557 73558 73558 73558 73558 73558 73559 73559	Modes the reases and somnia etc.)  Rows:  SE RXDDRUG RXDDRGILF 1 99999 1 INSULIN d00262 1 GABAPEN d03182 1 INSULIN G d04538 1 OLMESAR d04801 1 INSULIN A d04697 1 INSULIN G d04538	ON the	EY ARE F XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX	orescri	bed	FOR ( SC3 RXE Typ Res Typ Pur Typ Disc Pur	PRSD1 F  DRSD1 F  DRS	RXDRSD.  Detes migs synding on the petes migration of the peter migr	2 RXDRS ellitus rome ellitus ellitus wi erolemia ellitus ellitus	D3 RXDCOUNT 2 2 2 4 4 4 5 5 5 5 5 5

Key Variables: (which ones will be considered independent and dependent? Are you going to create new variables? What variables do you hypothesize beforehand to be most important?)

- Our goal is to model blood pressure (dependent variable) as recorded in the Demographics.csv dataset.
- We will be utilizing the data in demographics, diet, labs, examinations, and medications datasets (independent variables) to predict it (excluding the identifier, and the dependent variable).
  - Some examples of variables that we want investigate are:
    - Gender and age (and whether they match existing literature)
    - Alcohol and Caffeine,
    - Cholesterol levels, and
    - Drug Use.
  - We expect many variables to not be predictive (no silver bullet), but some may pop up and surprise us if we cast a wide net.
- We may create new variables to simplify the predictors and make them more meaningful.
  - There is a lot of missing data (e.g. questionnaire responses), and a lot of similar/ related predictors (e.g. do you feel pain, where do you feel pain, how long have you felt pain). To handle this, we may group up some of the existing variables, use indicators to capture the presence of it, or use dimensional reducing techniques (e.g. PCAs).

If we have time towards the end of the project, we may also try to model other health outcomes/ results to see whether we can predict them with accuracy or have additional interesting insights.

# **APPROACH/METHODOLOGY (8 points)**

Planned Approach (In paragraph(s), describe the approach you will take and what are the models you will try to use? Mention any data transformations that would need to happen. How do you plan to compare your models? How do you plan to train and optimize your model hyper-parameters?))

### EDA and Feature Selection

 First we plan on conducting exploratory data analysis on each of the datasets, assessing things such as sparsity, distributions, outliers. From data we deem okay to use, we will then run correlation tests to prune features and see how well raw data and engineered features relate to one another along with our response variable, blood pressure, and remove insignificant ones from the final feature list. Additional feature selection using RFE or model-specific ones might be leveraged as well.

### Model selection

 A suite of mostly linear and non-linear regression models such as linear-log, log-linear, log-log, and various polynomial models, other models such as XGBRegression may be explored as well, will be used to compare results, and the top one with the least RMSE will be chosen and used for our final model. Models might be compared using Cross-validation in order to prevent overfitting throughout our comparisons.

### Data Transformations

• We hope to experiment with both non-parametric models where distributions shouldn't matter too much, and parametric models as well, which might require log transformations, min-max scaling, and square root transformations in order to keep the data points on a relative scale to one another.. However other transformations such as one-hot encoding categorical variables, creating ratios of multiple numeric fields, and leveraging dimensionality reduction techniques such as PCA or K-mean clustering to reduce high-dimensional data sources may be used.

### Training and optimizing

• We plan to use a train validation and test split with ratios around 60%, 20% and 20% respectively with the training data used to enable the model to learn the feature weights, and the validation data used to tune the hyperparameters. Hyperparameter tuning will mainly be done manually with careful attention being paid to prevent over or under fitting of the model, and if time permits automatic methods such as gridsearch and hyperopt may be used Our test data will be held out and used to report the final performance of the model

Anticipated Conclusions/Hypothesis (what results do you expect, how will you approach lead you to determining the final conclusion of your analysis) Note: At the end of the project, you do not have to be correct or have acceptable accuracy, the purpose is to walk us through an analysis that gives the reader insight into the conclusion regarding your objective/problem statement

• Due to the high dimensionality of our data we assume many features looked at will not yield much predictive power. However there are many behaviors that medical science deems correlated to an individual having high blood pressure which we expect to see also represented in the model as feature weights/importances. Some of those features surround the use of tobacco and alcohol. We also expect overall diet and nutrition to play a big role with diets that veer off the recommended amounts in both nutrient breakdown and calories to impact blood pressure, both in terms of high and low depending on how excessive or restrictive each individual is.

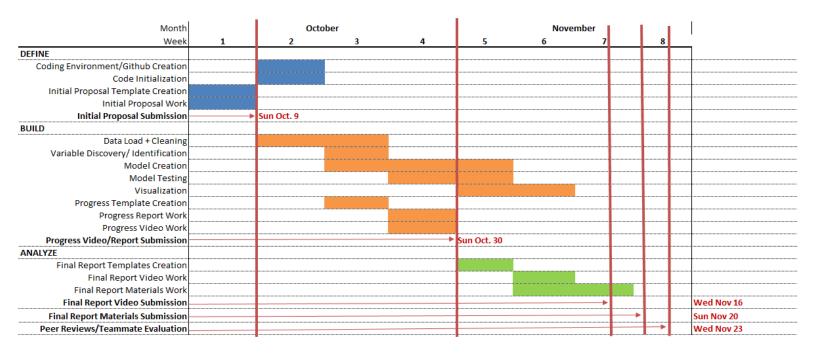
• The approach we take is going to be data-centric, we will not exclude anything based off intuition alone, however we will include and engineer features based off of our understanding of these scientifically proven correlations. In the end we hope to build as accurate of a model as possible, and by looking at the highly impactful features, give back to the audience some food for thought as to what factor inside and outside of their control may be involved with their blood pressure.

### What business decisions will be impacted by the results of your analysis? What could be some benefits?

Since our results will provide information on which factors could have a significant influence on a patient's blood pressure, the results can impact the health suggestions that are given to patients. We would like to understand what things a patient can and cannot change to improve health outcomes, which would be beneficial for healthcare providers and/or certain industries (such as food, medical, etc). Healthcare providers can determine treatments and the food industry can develop more suitable products that lead to a healthy diet, according to our analysis results.

Our analysis results can provide many benefits. First of all, this dataset prioritizes the behavior change of the patients by surveying diet information. It also has a broad selection of factors and considers the lowest hanging fruit, which could provide insights about what small changes can lead to big effects. The size of the dataset is also decent, which would give us a nice pool of data to use for analysis and prediction.

# PROJECT TIMELINE/PLANNING (2 points) Project Timeline/Mention key dates you hope to



#### achieve certain milestones by:

Table: General Project Timeline

As a general guideline, we are looking to have the following three general phases of our timeline done by these established dates:

1. DEFINE - Friday, October 14th

- 2. BUILD Friday, November 11th
- 3. ANALYZE Friday, November 18th

By establishing our project timeline with these deadlines in mind, we can not only reduce our overall workload as it pertains to this project, but also increase the quality of our final deliverables.

# Appendix (any preliminary figures or charts that you would like to include):

### **Additional Screenshots for Data**

				hic			_																																							
58QN 58 73557	DSRVYR RI	STATE RI	AGENDR R	DAGEYR RIDA	IGEMN RID	RETHS RIC	DRETHS R	DEXMON	RIDEXAGE	M DMQN	MILIZ DMQ/	ADFC DME	DBORN4 DI	MDCITZN E	MDYRSUS I	OMDEDUCE	MDEDUCD	VDMAR'	HDEXPRG :	HALANG	SIAPROXYS	LAINTRP	FIALANG F	APROXY	FIAINTRP N	HALANG	MIAPROX	MAINTRE	AIALANGED	MDERISE I	OMDFMSI D	MDHHSZE	IMDHH5Z D1	ADHHSZ D	MEHRGED	MDHRAC 69	DMDHRBS	DMDHREE	MOHRM.		INT2YF WTMEC2V 251.24 13451.04	SDMVPSU S	DMVSTR INDI	HHINZINDE		0.84
73558	8	- 2	1	54		4	-	- 1			1	- 1	- 1	- 1				- 4		- 1	- 2			- 2	- 4	- 1	- 2	- 1	- 1				0	- /	- 1	54	- 1				582.06 24471.77	- 1	108	-		1,78
73359		- 2	- 1	72				- 1						- 1						- :	- 4			- 4	- 4	- :	- 4		- 4		- 4		- 4		- 1	72	- 1		- 1		7214.8 57193.29	- 1	109	10		4.51
		- 4	- 1	72		3	3			119		-	- 1	- 1		-	- 4			- 1	- 4			- 4	- 4	- 1	- 4			- 4	- 4			- 4	- 1	72	- 1	- 1	- 1		201.18 55766.51	- 1	109	10		2.52
73560		- 4	- 1	73				- 1	- 1	119							-				- 1			- 2	- 4		- 2			4	4		- 4		-	33			-			- 4	116	15	15	2.52
73561 73562	8	2	2	56		3	3	1			2	-	- 1	1			5	1		- 1	2	- 2	1	- 2	- 2	- 1	- 2	- 2		- 2	- 2	0	0	- 2	- 1	/8	- 1	5	1		709.67 65541.87 978.14 25344.99	- 2	116	15		4.79
	8	2	1	56		1	1	1			1	2	1	- 1			4	3		- 1	2	- 2	2 1	- 2	- 2	- 1	- 2	- 2	1	1	1	0	0	0	1	56	1	- 4	3			1		,		4.79
73563	8	2	1		- 5	3	3	2		6			1	1						1	1	2	2 1	2	2					3	3	1	0		1	31	1	- 5	1		040.82 17913.56	1	105	15	15	- 5
73564	8	2	2	61		3	3	2			2		1	1			5	2		1	2	2	2 1	2	2	1	2	2	1	1	1	0	0	1	2	61	1	5	2		0325.1 61758.65	1	114	10	10	5
73565	8	1	1	42		2	2				2		1	1			3	1		- 1	2	2	2 1	2	2					4	4	1	1	0	1	42	1	3	1		307.68 0	2	106	15	15	5
73566	8	2	2	56		3	3	1			2		1	1			- 3	3		1	2	2	2 1	2	2	1	2	2	1	7	7	- 3	0	0	2	56	- 1	- 3	3		351.78 17480.12	1	112	- 4		0.48
73567	8	2	1	65		3	3	2			2		1	1			2	2		- 1	2	- 2	2 1	2	2	- 1	2	2	1	1	1	0	0	1	1	65	1	2	2		002.62 34795.43	2	112	- 3		1.2
73568	8	2	2	26		3	3	2			2			1			5	5	2		2	2	2 1	2	2	- 1	2	2	1	3	3	0	0	0	2	26	1	5	5		514.43 91523.52	2	113	15	15	5
73569	8	1	2	۰	10	5	7						1	1							1	2	2 1	2	2					4	4	2	0	0	2	30	1	4	1		73.504 0	1	110	77	77	
73570	8	2	2	9		5	7	1	- 10	109			1	1		2				- 1	1	2	2 1	2	2	1	2	2	1	4	4	0	3	0	2	43	1	5	3		913.82 12071,47	2	115	5		0.84
73571	8	2	1	75		3	3	1			2		1	1			5	1		1	2	2	2 1	2	2	1	2	2		2	2	0	0	2	1	76	1	5	1		988.47 46190.11	1	116	14	14	5
73572	8	2	2	10		4	4	1		125			1	1		3				1	1	2	2 1	2	2	1	2	2	1	3	3	0	2	0	2	30	1	- 4	- 5		689.38 12764.4	2	111	2		0.41
73573	8	2	1	10		4	4	1	12	128			1	1		4				1	1	2	2 1	2	2	1	2	2	1	5	5	1	2	0	1	44	1	3	1		03.213 6460.788	2	106	8		1.79
73574	8	2	2	33		5	6	1			2		2	1	4		5	1	2	1	2	2	2 1	2	2	1	2	2	1	4	4	1	1	0	2	33	2	5	1		155.31 15397.22	2	114	8		2.1
73575	8	2	1	1	18	4	4	1		19			1	1						1	1	2	2 1	2	2					2	2	1	0	0	2	19	1	2			74.951 7721.542	2	116	3		0.76
73576	8	2	1	16		4	4	2	15	94			1	1		9				1	2	2	2 1	2	2	1	2	2	1	6	6	2	2	0	2	35	1	4	77		433.75 12665.77	1	104	8		1.58
73577	8	2	1	32		1	1	1			2		2	2	4		1	6		2	2	1	1 2	2	1	2	2	2	2	7	5	3	2	0	2	32	2	3	6		284.65 39735.02	2	118	5		0.29
73578	8	2	1	18		1	1	1	21	18	2		1	1		15				1	2	2	2 1	2	2	1	2	2	1	7	7	0	2	2	2	60	2	3	1	1 11	592.85 11758.63	2	118	5		0.58
73579	8	2	2	12		3	3	2	34	47			1	1		6				1	1	2	2 1	2	2	1	2	2	1	4	4	0	2	0	2	38	1	5	1		248.99 70708.03	1	110	10	10	2.97
73580	8	2	2	38		4	4	1			2		1	1			3	1	2	1	2	2	2 1	2	2	1	2	2	1	6	6	0	3	0	2	38	1	3	1		388.92 27196.64	2	116	12	12	
73581		2	1	50		5	6	1			2		2	2	4		5	1		1	2	2	2 1	2	2	1	2	2	1	2	2	0	1	0	1	50	2	5	1	4 15	902.42 15754.25	2	106	15	15	5
73582	8	2	2	23		4	4	1			2		1	1			4	6	1	1	2	2	2 1	2	2					3	3	1	0	0	2	23	1	4	6	30	197.08 31181.89	1	108	3	3	0.69
73583	8	2	2	7		3	3	2	- 1	85			1	1		1				1	1	2	2 1	2	2					3	3	0	1	0	1	44	1	5	1	5 47.	574.65 48196.41	1	110	15	15	5
73364	5	2	1	13		3	3	2	22	36			1	1		7				1	1	2	2 2	2	2	1	2	2	1	3	3	0	1	0	2	35	1	4	1	72	700.36 72162.24	1	105	,	2	3.07
73585	8	2	1	28		5	6	2			1	1	2	1	6		4	1		1	2	2	2 1	2	2	1	2	2	1	2	2	0	0	0	1	28	2	4	1		063.14 18590.97	1	110	7		2.26
73586	8	2	2	4		5	6	1		58			1	1						1	1	2	1	2	2					4	4	2	0	0	1	33	2	4	1	4 11	635.24 13569.68	1	116	6	6	1.27
73587	8	2	1	14		5	6	1	12	34			1	1		8				1	1	2	1	2	2	1	2	2	1	7	7	1	1	0	1	32	1	3	4	16	220.74 15523.09	2	115	15	15	3.33
73588	8	2	1	11		4	4	1	14	40			1	1		4				1	1	2	1	2	2	1	2	2	1	3	3	0	2	0	2	31	1	- 4	5	84	24,955 8291,079	1	109	6	6	1.54
73589	8	2	1	35		3	3	2			2		1	1			3	3		1	2	2	2 1	2	2	1	2	2	1	1	1	0	0	0	1	35	1	3	3	57.	197.64 61374.1	1	114	5	5	1.74
73590	8	2	1	2		1	1	1		30			1	1						1	1	2	1	2	2					5	5	3	0	0	1	24	1	4	1	4 12	254.74 12849.22	1	106	5	5	0.8
73591	8	2	1	6		3	3	1		76			1	1		0				1	1	2	1	2	2					4	4	0	2	0	2	40	1	- 5	1		374.79 51133.13	2	114	15 A		
73592	8	2	1	29		1	1	2			2		1	1			4	5		1	2	2	2 1	2	2	1	2	2	1	3	1	0	0	0	1	24			5	64	083.31 68450.78	2	107	12	6	2.14

### Diet.csv

16888.33 12930.89	1	49	2						13											1574	43.63	235.59	176.47	10.8	52.81	17.819	18/493	8.825	209	3.92								430	1.057
17932.14 12684.15	1	50	2	2	4	1	1	1	13	1	2	3	1	1	2					5062	335.13	423.70	44.22	16.7	124.29	53.400	35,481	20.505	2564	20.04	0	3477	1455	0	145	12	0	600	4.176
9641.81 39394.24	1	49	2	2	18	6	1	1	13	1	1	2	2		1		7		27	1743	64.61	224.39	102.9	9.9	65.97	25.263	20.902	12.953	88	8.87	0	616	655	21	449	35	515	300	1,648
142203.1 125956.4	1	54	2	2	21	2	1	1	12	1	1	2	2		2				22	1490	77.75	162.92	50.55	10.6	58.27	22.511	15,985	9.617	197	2.15	0	515	545	20	227	223	0	583	1,407
6052.36 39004.89	1	63	2	2	18	1	1	1	13	4		4	2		1			91	7	1421	55.24	178.2	87.78	12.3	55.36	4.479	26,216	1.263	41	53.17	0	738	863	0 0	1476	0	0	0	1,484
0 29.002	1	49	2	1	11	2	1	1	11	1	2	2	1	1	2				27	1785	55.11	189.59	83.75	22.6	92.92	22.155	40.013	23.55	534	12.3	0	272	355	172	776 1	200 2	2229	2101	1.227
1417.22 40785.78	4	54	1	2	2	8	1	2	13	4		1	2		2				30																				
78988.76 52178.16	1	54	2	2	12	7	1	1	13	4		1	2		2				21	2585	91.15	300.16	129.23	24.8	93.03	25.989	20.97	37.652	152	16.11	7.43	863	1372	1970 3	5153	55 3	1030	843	2,201
0697.83	1	49	2	1	3	2	1	1	13	4		3	2		2				34	1580	42.26	226.32	85,92	15.1	41.31	39.911	14,878	15.6	39	5.75	0	478	491	0	182	0	0	211	1.431
M500.04 0	1	61	2	1	16	7	1	1	13	1	1	4	2		2				7	2021	38.09	216.59	99.01	7	82.55	28.438	24.3	24,586	81	5.89	0	9	11	3	19	8	0	55	1,461
6404.07 43555.55	1	87	2	2	12	7	1	1	13	1	2	3	2		2				30	3145	139.21	227.63	83.91	37.6	172.21	66.642	57,929	32.571	625	17.45	0	868	1427	1951 8	5699	134 13	3057	1447	1.877
1176.25 20410.41	1	22	2	2	15	2	1	2	1	1	1	3	2		2				15	2220	76.4	299.40	146.21	10.7	01.19	25.657	26.002	17.215	406	6.19	0	524	556	21	339	78 6	3040	416	1.616
65076.73 39020.99	1	25	2	2	12	7	1	1	13	1	2	3	2		2				34	1076	33.4	150.93	61.42	14.9	39.68	34.153	14.336	8.762	80	1.98	0	231	607	52 6	4463	57		201	1.693
2965.83 25811.62	1	61	2	2	8	4	1	1	2	1	1	4	2		2				22	1545	20.65	178.12	115.01	5.8	79.54	23.059	18.599	35.671	97	7.05	0	124	677	0 0	6641	1	0	28	0.578
0	5																																						
14293 10295.52	1	49	2	2	18	1	1	1	11	4		4	2		2				35	5621	274.72	575.61	215.83	36	249.49	93.879	92,405	50.794	687	23.81	0	605	839	115 2	2711	13 2	2827	2483	4.651
057,793 0	1	os	2	1	16	7	1	3	13	4		3	2		2				31	888	21.6	124.92	79.45	5.9	37.08	13.026	9.99	3,333	49	2.52	0.05	190	200	1	50	0 1	058	109	0.350
756.102 0	1	49	2	1	23	6	1	1	13	4		1	2		2				6	1012	48.91	345.46	44,67	4	25.14	7.143	8.835	6.313	111	1.45	0	2	2	0	1	1	0	9	1.081
2434.51 0	1	49	2	1	13	1	2	1	13	1	2	4	2		2				22	3354	144.52	385.05	158.45	33.4	121.74	45.435	42,117	16.147	561	6.04	0	254	285	10	377	3 3	3562	246	1,863
26314.1 44781.19	1	49	2	2	35	4	1	1	13	1	1	4	2		2				7	555	81.61	113.89	5.31	15.6	18.24	6.431	4.534	3.264	170	1.62	0	92	93	0	16	0	0	75	1.125
6016.59 43130.65	1	50	2	2	34	1	1	1	13	1		3	2		2				25	1711	81.54	211.42	82.71	11	60.8	25.869	13,464	11.279	149	4.42	0	624	651	1	351	0 3	1796	187	1,060
6109.18 0	1	54	2	1	30	4	1	1	13	1	1	4	2		2				18	2421	87.39	305.69	153.65	14.3	97.72	26.594	37,977	25.078	256	6.98	0	91	116	39	253	21	515	523	1.503
1269.094 27439.89	1	25	2	2	22	7	1	1	13	4		4	2		2				27	3053	96.42	534.92	161.91	50.3	63.55	20.273	22.895	13.006	66	8.2	0	320	552	1358 2	2153	22	421	274	2,460
0	5																																						
8522.69 29514.52	1	49	2	2	26	7	1	2	1	1	1	3	2		2				27	571	25.81	94.6	41.03	11.8	12.2	2.644	3.692	4.153	40	2.5	0	243	262	22	419	106 3	1684	294	0.478
8109.25 44089.08	1	54	2	2	20	1	1	1	13	4		2	2		2				7	663	13.11	87.44	29.45	9.6	29.86	7.801	7.423	12.321	17	3.13	a	113	165	121	566	0 3	1088	287	0.802
1674,87 0	1	39	2	1	10	4	1	1	13	1	3	1	1	1	1	1				2515	285.83	175.42	52,01	45.7	80.46	21.50	27,719	20,488	775	17.6	0	136	6541	430 73	3940	9	0 1	10020	2,000
888.037 5138.003	1	25	2	2	22	7	1	2	13	4		4	2		2				12	800	24.52	93.32	45.33	8	37.64	12.377	16.628	5.88	90	2.56	0	10	122	14	1326	35	3	279	0.69
523,513 6344,872	1	71	2	2	48	6	1	1	13	1	2	3	2		2					875	23.33	349.99	55.22		21.22	5.325	2,373	3,511	43	0.59	0	146	175	23	294	20	1	157	0,654
615.804 7657.103	1	59	2	2	8	1	1	1	11	4		1	2		2				17	2226	51.48	348.91	129.33	19.3	70.9	28-214	16.228	11.943	185	6.18	0	405	442	79	331	94	5	1441	1,285
7969.85 29589.81	1	25	2	2	23	6	1	1	13	1	2	3	2		2				23	1457	70.54	232.55	137.22	15.4	31.66	22.229	11.98	6.145	153	2.95	0	562	572	17	102	50	637	165	1.500
058.356 6315.401	1	49	2	2	27	6	1	2	13	4		4	2		2				12	1041	48.28	107.46	79.63	2.2	46.94	16.33	18.1	8.022	168	4.2	0.08	186	190	0	25	0	0	83	0.423
18680.16 21973.57	1	25	2	2	32	6	1	2	1	1	2	4	2		2				22	1100	46.04	349.60	61.62	12.9	59.02	24.779	17.059	11.143	119	4.3	0	448	461	1	174	2 2	2037	106	0.906
140175.1 216982	1	25	2	2	42	5	1	1	13	1	- 1	2	2		2				22	3655	106.97	414.17	285.23	8.5	177.38	34.358	42.087	83.996	441	11.23	0	381	824	803	5127	21 3	1122	4504	1.08
0362.89 166815.1	1	63	2	2	24	6	1	2	11	4		2	2		2				21	1803	65.43	289.17	154.83	21.8	48.72	18.183	16,551	8.88	292	4.68	0	449	506	150	455	354 4	4472	913	1,439
9001.11 144981.4		61							12										33	1550	41.45	222.17	80.82	27.9	41.89	33.985	14.688	11.714	199	8.68								407	0.856

## Labs.csv

7	4.3	4.8	39	3447.6	11.03				4.1	41	129	16	26	10	3,57	27	9.5	2,375	168 4.	544	174	97 1	21 106.5	99 2.4	21	554	50,75	10	58	10.4	4.3	118	136	250	4.8	1.388	0.8	15.68	6.5	65	140	1.581	3.3	195.3	4.7	42.2	2.2	42.5	5
3	153	153	50	4420	306				4.7	47	97	13	29	16	5.71	23	9.2	2.3	167 4	119	147	98 0	79 69.1	84 3.1	31	219	12.16	15	79	14.1	4.3	122	128	265	3.9	1.259	0.9	15.39	7.8	78	257	2,902	4.7	279.6	12.6	27.3	7.6	58.4	4
,	11.9	11.9	113	9989.2	10.53	42196.9	57	0.57	3.7	37	99	22	16	14	5	23	8.9	2.225	127 3.	284	44	105 1	22 107.1	85 2.7	27	183	10.16	13	58	17.6	4.2	134	142	288	4.2	1.356	0.6	10.26	5.9	59	51	0.576	5.7	339	7.2	13.9	11.5	63.2	2
	16	16	76	6719.4	21.05																																								7.0	29.6	9.2	59.1	1
	255	255	147	12994.5	173.47	142255	92	0.92	4.3	43	75	35	25	31	11.07	31	10	2.5	207 5.	153	60	200 0	73 64.5	55 2.0	21	204	5.77	32	91	16.3	4.1	201	142	250	4.4	1.421	0.5	5.55	7.1	71	55	0.994	4.2	245.5	6.6	20.5	6.9	65.7	7
2	123	123	74	6541.6	166.22				4.3	43	95	24	16	18	6.43	25	9.3	2.325	230 5.	148	340	103 0	89 78.1	68 5	30	104	5.77	22	52	9.3	3.3	211	143	287	3.3	1.066	0.5	8.55	7.8	73	327	3.692	9.1	541.3	9.4	17.4	6.6	69.2	2
3																																																	
ŧ.	19	19	242	21392.8	7.85	34054.1	77	0.77	3.9	39	72	20	21	17	6.07	25	9.9	2.475	167 4.	819	120	307 0	92 81.3	83 2.5	25	107	5.94	17	71	12.7	8.9	133	140	281	3.6	1.162	0.6	10.26	6.4	64	68	0.768	5.1	803.3	5.2	30.2	5.9	58.8	3
5	1.3	1.3	10	1591.2	7.22				4.1	41	93	23	24	9	3.21	26	9.5	2.375	270 7.	189	90	203 0	55 48.6	62 2.5	25	100	6	21	36	6.4	4	132	139	277	5.2	1.679	0.4	6.84	6.6	66	262	2.950	3.5	200.2	9.5	31.3	7.7	57.5	5
	35	35	215	19006	16.28				4	40	67	29	20	35	5.36	26		2.175	170 4.	196	220	305 0	97 85.7	75 3.3	33	104	5.77	24	83	14.5	4.8	168	140	281	3.3	1.005		11.97	7.5	75	39	0.44		374.7	6.3	22	9.4	64	4
ŝ	25	25	31	2740.4	80.65	16002.5	50	0.50	4.5	45	65	20	23	12	4.28	24	9.3	2.325	174	4.5	60	206 0	74 65.4	42 2.5	25	81	4.5	13	82	14.7	4	132	139	276	4.6	1.485	0.6	10.26	7	70	31	0.35	4.5	267.7	6	20.9	7.7	70.1	1
9	27.2	27.2			23.45																																												
1	25.8	25.8			14.58				4.3	43	63	28	27	17	6.07	24	9.3	2.325	157 4	.05	108	103 1	19 105	2.7	27	126	6.99	24	65	11.6	4.8	112	138	279	4.4	1.421	0.7	11.97	7	70	284	3.206	6.5	386.6	5.4	30.2	9.3		
2	51.	51			35.42																																								7	32.9	6.5	50	,
	6.3	6.2			8.29																																												
1	13	13	173	15255.2	7.51	0	61	0.51	4.3	43	35	20	20	11	3.53	23	9.1	2.275	130 3.	162	100	200 0	59 52.	16 2.5	25	55	4.54	21	35	6.5	4.5	120	139	276	2.7	0.872	0.9	15.39	7.2	72	23	0.25	2.7	160.6	4.6	25.6		65.2	
5																																													6.1	64.1		22.2	
	5	5		14574,4		3038.13	55	0.55	5.1	51	110	25	15	14	5	27		2.475	157 4	.06			89 78.	68 2.1	. 2	83	4.61	15	49	8.8	4.7	135	243	285	4.1	1.324	0.4	6.84	7.2	72		0.553	5.4		5.1	35		53.2	
7	17	17		16884.4		5784.91	111	1.11	4.5	45	76	90	44	17	6.07	26	9.3	2.325	185 4	.81	146	304 0	68 60.3	11 2.5	25	97	5.88	90	77	13.8	8.9	114	139	279	2.8	0.904	0.7	11.97	7.4	74	151	1.705	6.1	852.8	8	40.6	9.2	45.7	/
	10.6	10.6		17769.4		0																																											
,	1.2	1.2		1979.0	0.35				4.7	47	198	20	35	10	3.57	25	9.9		136 3.		73	303 0	57 50.2		30	102	5.66	10	55	9.9	4.3	122	140	279	5.1	1.647	0.2	3.42	7.7	77	63	0.711		184.4	6.3	32.7	10.3	53.8	
•	3.6	3.6		11403.6		6414.04	110	1.1	3.8	38	52	17	23	10	3.57	25	8.9	2.225	228 5.		61	207 0	72 63.0		2	94	5.22	19	96	17.2	3.5	124	140	278	3.9	1.259	0.5	8.55	6.7	67	64	0.723		255.8	5.6	30.8		49.3	
	10.2	10.2		18210.4		7181.38	90	0.9	4.3	43	60	26	20	11	3.93	26	9.2	2.3	203 5	.25	215	204 0	55 83.5	98 3.1	31	104	5.77	20	126	22.6	4	158	139	277	3.5	1.13	0.5	8.55	7.4	74	107	1.208	5.3	315.2	4.4	31.6	7.3	60.7	4
	10.8	10.8		12552.8																																													
	29	29		13349.4					4.2	42	297																																						
4		2.7			4.72				4.2	42	297	19	34	21	5	23	9.7	2.425	195 5.	342	96	105 0	21 106.5	16 2.1	- 21	91	5.05	11	100	17.9	3.7	159	143	256	5.5	1.908	0.5	15.35	6.5	65	241	1.784	2.8	226 392.6	7.8	29.6	9.2	49.4	
5	2.7	2.1	100	8890	2.7				4.7	47	13	21	25	- 23	8.21	20	7.0	24	193 4.	191	196	200 1	21 106.5	20 1.4		- 80	4.72	15	- 00	15/4	3.7	110	243	200	4.0	1.55	0.3	15.35	0.5	60	158	1.701	6.6	332.0	6.8	31.3	8.8	50.5	-
	9.2	0.0	***	20362.4	20.	2066.13		0.84	4.5	45	120	47		10	5.71	24	9.5	1.175	159 4.		255	105 0	88 77.°	79 2.5	21	- 00	5.11	32	67	15.6	3.9	132	139	278	4.4	4.401		17.1	-	20	83	0.937	7.1	422.2	10.8	24.6	21		
		4.1		13435.8		V060.13	64	V.04	+.5	45	129	37	53	16	5.71	24	2.3	43/3	139 4.	LLC	433	200 0	60 71.	19 23	- 2	92	5.11	32	67	15.6	3.3	194	233	418	4.4	1.421	- 1	1/.1	7	N	03	0.92/	7.1	444.3	20.8	32.8	7.1	52.4	-
		11.4		17500					4.4			22	21	12	4.20	22	9.2	2.3	195 5	200	121	105 0	ED 77	70 .			5.00	17		16.5	4.2	124	140	279	3.1	1.001	0.0	13.65	7.6	71		0.734		401.5	- :	20.0		57.8	
		4	200	17500	5.7				***	44	/0	22	- 21	12	4.20	43	5.2	4.4	200 5.	~	***					36	5.44	17	32	20.3	4.2		200	2.79	-1	1.001	v.8	44.00	1.4		40	0.700	0.8	4,42		40.0		37.8	+
			- 00	2222.0	15.63																																									23.3	2.0		
	,	,			6.3							28			2.5																															29.8			

### Questionnaire.csv

SEQN A	0011A	ACD0118	ACD011C	ACD046	ACD1	10 ALO	101 4	ALQ110	ALQ120	Q ALQ12	OU ALC	1130 A	4LQ141Q	ALQ141	U ALQI	151 ALC	0160 8	PQ620	BPQ090	BP0035	BPQ0464	BPQ050	EPQ056	5 BPD05	8 BPC	059 8	Q060	800,060	BPQ070	BPQ	900 BPQ	1000 CB	0070 CI	BD090	280110	C80120	CB0130	0 HSD0	110 HSQ	500 HSI	1510 H	Q520 H	SQ571	H5Q580	HSQ590 F	SAQUEX C	iq010 (c	iquu cs	Q030 r	CSQ640	CSQ060	CSQ070	CSQ380	CSQ0504	A CSQ050	3 (CSQ090C )
73557	1						1			1	3	1	0			1		1	1	62		1	2	2		2	1			2	1	1	300	0	50		0	85	2	2	2	2	2		2	2	2	3	2	2				2	3	3 3
79558	1						1			7	1	4	2		1	1	0	1	1	53		2		1	52	2	1			3	1	1	642	214	128		40	20	4	2	2	2	2		2	2	2	3	2	2			- 7	2	3	3 3
71550	1						1			0						2		1	1	40		1	1	2		2	1			1	1	1	150	25			00	0	3	2	2	2	2		2	2	2	3	2	2			7	2	3	3 3
73560	1																																400	0	50		50	30		2	2	2				1										
71561	- 1						1			0						2		1	2	55		1	1	1	22	2	2	-		1	2		200	0	40		0	0	5	2	2	2	2		2	2	2	3	1	2				2	3	3 3
79562					4		1			5	3	1	0			2	0	1	1	35		1	1	1	365	1	1			2	1	1	150	60			60	0	5	2	2	2	2		2	2	1	2	1	1	- 4	( )	2 3	1	2	2 2
73563																																	642	50		4	00	0																		
73564	1						2		1	2	3	1	9			2		1	1	47		1	2	2		2	2			2	2		400	100	90	2	00	0	3	2	2	2	2		2	2	2	2	1	2	- 3	( 7	2 7	2	3	3 3
71565					5													2						2		2	2	-		3	2		500	0	60	3	00	40									2	3	2	2			- 2	2	3	3 3
73566	1						1			1	1	1	0			2		2						2		2	2			2	2		1000	200	200		50	25	3	2	2	2	2		2	2	2	3	2	2			- 2	2	3	3 3
73567	1						1			4	1	3	0			2		2						2		2	2			4	2		125	25			0	0	3	1	2	2	2		2	2	2	3	1	2			2	2	3	3 3
73568	1						1			2	1	2	1		3	2	0	2						2		2	2						662	30			50	0	1	2	2	2	2		2	2										
73569																																	1000	0		10		0																		
73570																																	400	0	214	2	00	0		2	2	2				1										
79571	1						2		1	2	8	1	0			2		1	1	60		1	1	1	12	1	1			1	1	1	240	0		2	00	0	3	2	2	2	2		1	2	2	8	1	2			2	2	3	3 3
71572	1																																200	200	50		50	0		2	2	2				1										
79578	1																																300	0	500		75	75								1										
72574						1	2		2									2						1	1	2	2						500	100		5	00	0	1	1	1	1	2		1	2										
75575																																	100	30	30		25	40		2	2	2				1										
73576	1																	2						2		2	2	_		2	2		1000	0		2	00 1	100	1	2	2	2	2		2	2										
73577					1		1			20	3	15	10		3	2	0	2						2		2	2	_					500	200	200	1	50 ]	100	3	2	1	2	2		2	2										
73578					2		2		2									2						2		- 2	2	_		2	2		900	0			0	0	2	2	2	2	2		1	2										
73579	1																																300	0	150	- 4	00	0	2	2	2	2				2										
79580	- 1						1			1	3	3	0			1		2						2		2	2	_					500	0			30	0	2	2	2	2	2		2	2										
73581						1	2		1	0						2		2						1	52	1	1			1	2		500	0		2	00	0	2	2	2	2	2		2	2	2	3	2	2			2	2	3	3 3
79582	1																	2						2		2	2			2	2		126	0			25	36								2										
73503	- 1																																600	100		2		20		2	2	2				1										
73584	1																																600	0	30	3	00	0	3	2	2	2				2										
71565						5	1			4	3	4	0			2		2						1	2	2	2	_		4	2		642	0			10	0	1	2	2	2	2			2										
73586						5																											400	0			0	0		2	2	2				1										

# Examination.csv

SEQN P	ASCST1 PE	ASCTM: PEASCCT1	BPXCHR BPAA	ARM BPACS	Z GPXPL	S DPXPUL	SEPXPTY	( BPXXVI	LI BPXSY	ri BPXD	1 BPAD	N1 BPXS	5Y2 8PX0	N2 BPAEN	2 BPXS	Y3 BPXDI	3 EPAENS	DPX51	r4 BPXD	64   BP	MEN4 DAY	IDSTAT: DM		BMOXRECU	BMIRECUN BMXHEAD BMIHEAD I	THIMS THOMS	BMXBMI BMD	BMIC BMXLEG BMILE	S DMXARM(BMIA	IFIME BMXAFFM(BMIAR	IMC BMXWAIS BMFWA			IXSAD3 BMXSAD	4 BMDAYSA BMDSA	DC MGDEXST; M	/GD050 MGD060
73357	1	620		1	4	85	1		140	122	72	2	114	76	2	102	74	2				1	78.3			171.3	25.7	39.2	40.2	35.3	100		20.6		20.6	1	2
73558	1	766		1	4	74	1	1	170	156	62	2	160	80	2	156	42	2				1	89.5			176.8	28.6	4)	41.5	34.7	107.6	24.2	24.5		24.4	1	2
73559	1	665		1	4	68	1	1	160	140	90	2	140	76	2	146	80	2				1	88.9			175.3	28.9	40	41	33.5	109.2	25.8	25.4		25.6	1	2
73560	1	802		1	2	64	1	1	120	108	28	2	102	34	2	104	22	2				1	32.2			137.3	17.1	2 22.5	29.5	21	61	14.8	15		14.9	1	
73561	1	949		1	3	92	1	1	170	135	56	2	134	55	1	142	55	2				3	52			162.4	19.7	35.3	37.5	25.2		1				1 1	2
73562	1	1054		1	5	60	1	1 :	180	160	84	2	158	82	2	154	80	2				1	105			158.7	41.7	84.2	36.2	41.8	129.1	29.1	29.7	29 29.3	29.1	1	1
73563	1	90	152				1															1	7.4	68	43.7				13.7	14.9							
73564	1	954		1	5	82	1	1	150	118	80	2	124	80	2	126	82	2				1	98.4			161.8	85.7	87.1	39.3	88	110.8	26.6	25.7		26.7	1	2
73566	1	625		1	4	86	1	1	140	128	74	2	124	72	2	114	72	2					61.8			152.8	26.5	32.4	33.5	29	85.5		19.8		19.9	1	2
73567	1	932		1	3	70	1	1	170	140	78	2	142	78	2	142	76	2					65.3			172.4	22	40	40.3	27.5	99.7		20.1		20	1	2
73568	1	585		1	3	70	1	1	120	106	60	2	100	62	2	108	66	2				1	47.1			152.5	20.3	34.4	32.6	25.8	73.7	14.4	14.6		14.5	1	2
73570	1	710		1	2	75	1	1	150	102	44	2	104	42	2	102	60	2				1	31.9			135.5	17.4	2 30.4	28.7	20.8	65.6	15.3	15.5		15.4	1	
73571	1	1193		1	4	72	1	1	150	124	68	2			2	128	68	2	128	64	2	1	102.4			172.5	34,4	35.5	38	36.5	122.1		29.7		29.7	1	2
73572	1	557		1	3	80	1	1	110	83	54	2	88	54	2	94	58	2				1	41.7			145.4	19.7	2 36.8	31.3	24.7	66.4	15.2	15.4		15.3	1	
73573	1	599		1	3	74	1	1	120	94	62	2	94	60	2	92	54	2				1	50.1			158.38	20	3 39.5	34.8	25.7	72.5	15.3	15		15.2	1	
73574	1	710		1	3	23	1	1 :	160	122	56	2	124	66	2	115	64	2					56.8			150	22.0	37.5	23.4	26.5	70.7	15.1	15		15.1	1	2
73575	1	291	110				1															1	9.4	80					16	13.9							
73576	1	584		1	4	62	1	1 :	130	108	68	2	110	62	2	110	70	2				1	67.3			170.4	23.2	2 42	37	30.7	74.2	17	17		17	1	
73577	1	750		1	4	56	1	1	140	118	76	2	118	74	2	122	78	2				1	79.7			166-2	28.9	35.5	37	34.5	100	22.6	23		22.8	1	2
73578	1	607		1	5	68	1	1	140	120	58	2	124	64	2	124	70	2				1	109.4			175.2	35.6	4 44.1	35.5	38.2	110.2	25.6	25.9		25.8	1	
73579	1	648		1	3	96	1	1	130	108	72	2	116	70	1	110	99	2				1	40.2			161	15.5	2 35.7	34.4	21.8	62.6	13.6	14.1		14	1	
73580	1	805		1	5	70	1	1 :	140	124	84	2	118	66	1			2	114	58	1	1	93.5			161.4	35.9	40	19.5	41	107.4	26.5			26.5	1	2
73581	1	699		1	4	74	1	1 :	160	138	80	2	132	80	2	130	76	2				1	80.9			185	23.6	42.8	38	30.9	99.3	22.5	22.1		22.3	1	2
73582	1	527		1	3	88	1	1	120	98	56	2	96	48	2	98	52	2				2	46.7			158.3	18.6		1	1	1	1				3	
73583	1	283	94				1															1	19.7			116.9	14.4	2	23.7	17.7	49.9					1	
73584	1	753		1	3	94	1			108	54	2	110	52	2	110	54	2					58-1			144.9	25.3	3 37.5	37	28.8	77		16.2		16.4	1	
73585	1	623		1	5	72	1	1 :	120	105	70	2	106	74	2	100	66	2					92.2			175.1	30.1	40.5	40	42.6	90.3	10.3	19.3		18.3	1	2
73386	1	318	90				1																18.6			109.4	13.3	2	21.3	10.3	53.7						
73587	1	930		1	5	74	1	1 :	140	112	72	2	110	68	1	114	66	1				1	110.2			168.8	38.7	4 45.8	38.6	43.4	115.9	25			25	1	
73588	1	581		1	2	76	1	1	120	100	42	2	98	0	2	100	0	2				1	31.5			141.5	15.7	2 35	31.4	20	56		13.8		13.8	1	
73589	1	1018		1	4	88	1	1	150	124	78	2	118	66	2	118	72	2				1	78.3			172.9	26.2	39.1	37.2	81	94.6	21.2	21.3		21.3	1	2
73599	1	304	204				1															1	15.2	99.5		98.3	15.7	2	20	16.4	46.4						
78591	1	281	304				1																21.6			112.2	17.2	3	22.8	19.1	54.7					1	
73592	1	619		1	4	86	1	1	140	118	86	2	120	78	2	122	76	2				1	92.9			173.1	31	40.8	40.1	32.5	107.7	25.5	25.5		25.5	1	2

# Medications.csv

QN	RXDUSE	RXDDRUG	RXDDRGII	RXQSEEN	RXDDAYS	RXDRSC1	RXDRSC2	RXDRSC3	RXDRSD1	RXDRSD2	RXDRSD3	RXDC
73557	1	99999										
73557	1	INSULIN	d00262	2	1460	E11			Type 2 dia	abetes mel	litus	
73558	1	GABAPEN	d03182	1	243	G25.81			Restless I	egs syndro	me	
73558	1	INSULIN G	d04538	1	365	E11			Type 2 dia	abetes mel	litus	
73558	1	OLMESAR'	d04801	1	14	E11.2			Type 2 dia	abetes mel	litus with	
73558	1	SIMVASTA	d00746	1	61	E78.0			Pure hype	ercholeste	rolemia	
73559	1	INSULIN A	d04697	1	365	E11			Type 2 dia	abetes mel	litus	
73559	1	INSULIN G	d04538	1	4380	E11			Type 2 dia	abetes mel	litus	
73559	1	PANCRELI	d01002	1	365	K86.9			Disease o	f pancreas	, unspecifi	•
73559	1	SIMVASTA	d00746	1	2920	E78.0			Pure hype	ercholeste	rolemia	
73559	1	VALSARTA	d04113	1	3650	I10			Essential	(primary) h	ypertensi	•
73560	2											
73561	1	55555		1	152	55555						
73561	1	CARVEDIL	d03847	1	6205	110			Essential	(primary) h	nypertensi	•
73561	1	LEVOTHYR	d00278	1	6205	E03.9			Hypothyr	oidism, un	specified	
73561	1	VALSARTA	d04113	1	4380	110			Essential	(primary) h	nypertensi	•
73562	1	AMLODIPI	d00689	2	365	110			Essential	(primary) h	nypertensi	C
73562	1	ATORVAS <sup>*</sup>	d04105	2	1825	E78.0			Pure hype	ercholeste	rolemia	
73562	1	LISINOPRI	d00732	2	3650	110			Essential	(primary) h	nypertensi	c
73562	1	MINOXIDI	d00135	2	365	110			Essential	(primary) h	nypertensi	C
73562	1	NAPROXE	d00019	2	730	55555						
73562	1	PAROXETI	d03157	2	5475	G47.0			Insomnia			
73562	1	PRASUGRE	d07409	2	365	Z95.5			Presence	of coronar	y angiopla	9
73563	2											
73564	1	CITALOPR	d04332	1	1825	F32.9			Major de	oressive di	sorder, sin	1
73564	1	LORAZEPA	d00149	1	5475	F41.9			Anxiety d	isorder, ur	specified	
73564	1	ZOLPIDEM	d00910	1	1825	G47.9			Sleep dis	order, unsp	pecified	
73565	1	BUPROPIC	d00181	2	5	F17.203			Nicotine	dependen	ce unspeci	f
73566	1	ACETAMIN	d03428	1	2555	M79.1			Myalgia			
73566	1	BACLOFEN	d00967	1	2555	M62.83			Muscle sp	asm		
73566	1	GABAPEN	d03182	1	2555	M62.83			Muscle sp	asm		
73567	2											
73568	2											
73569	2											
73570	2											
73571	1	ESOMEPRA	d04749	2	3650	K21			Gastro-es	ophageal r	eflux dise	