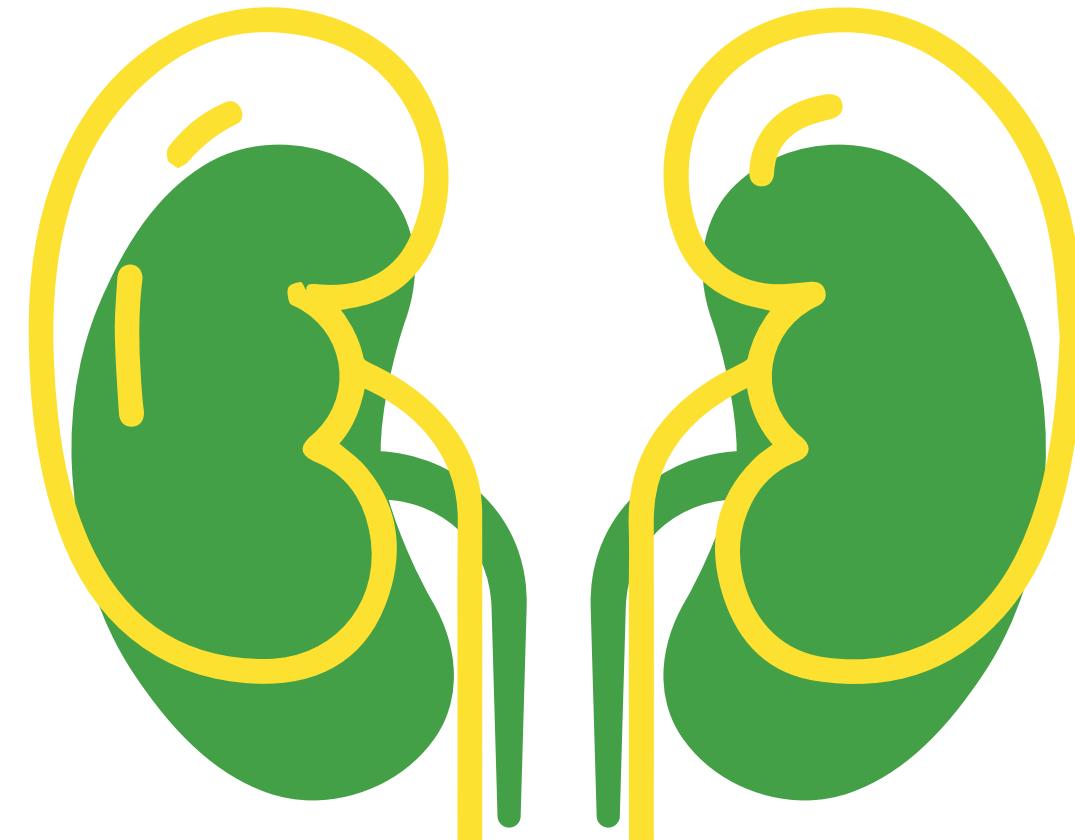
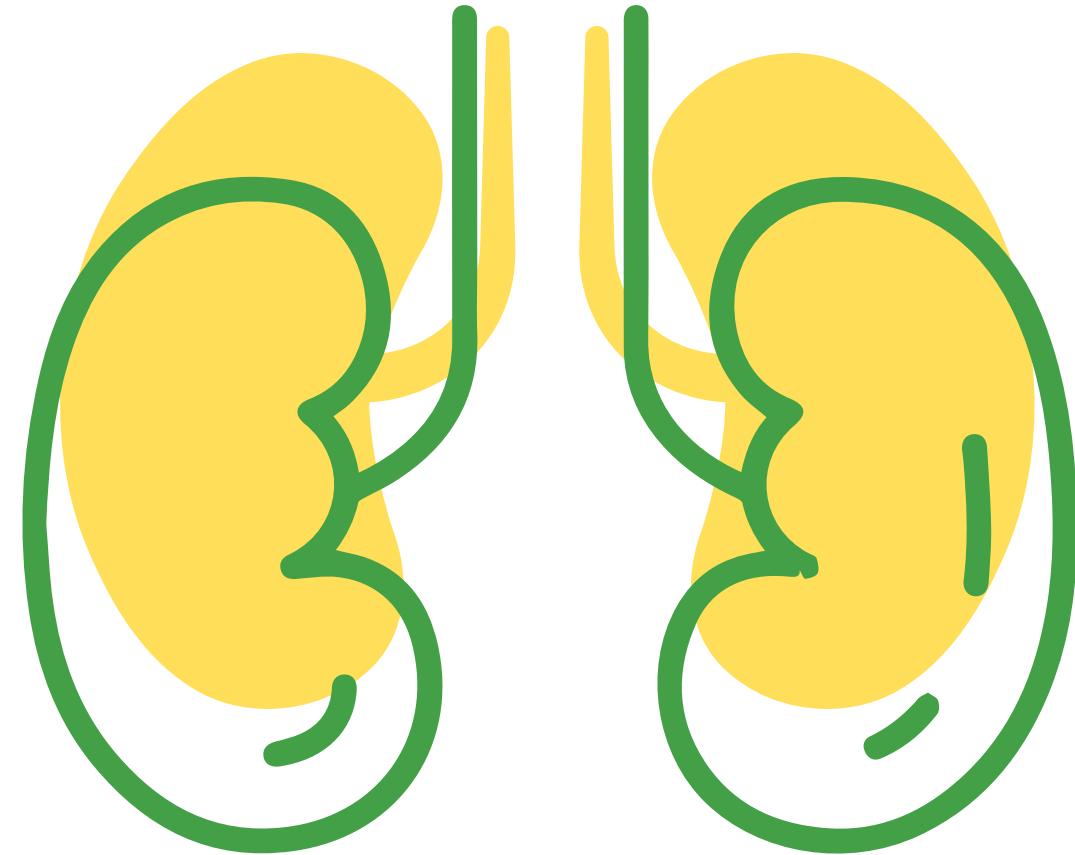


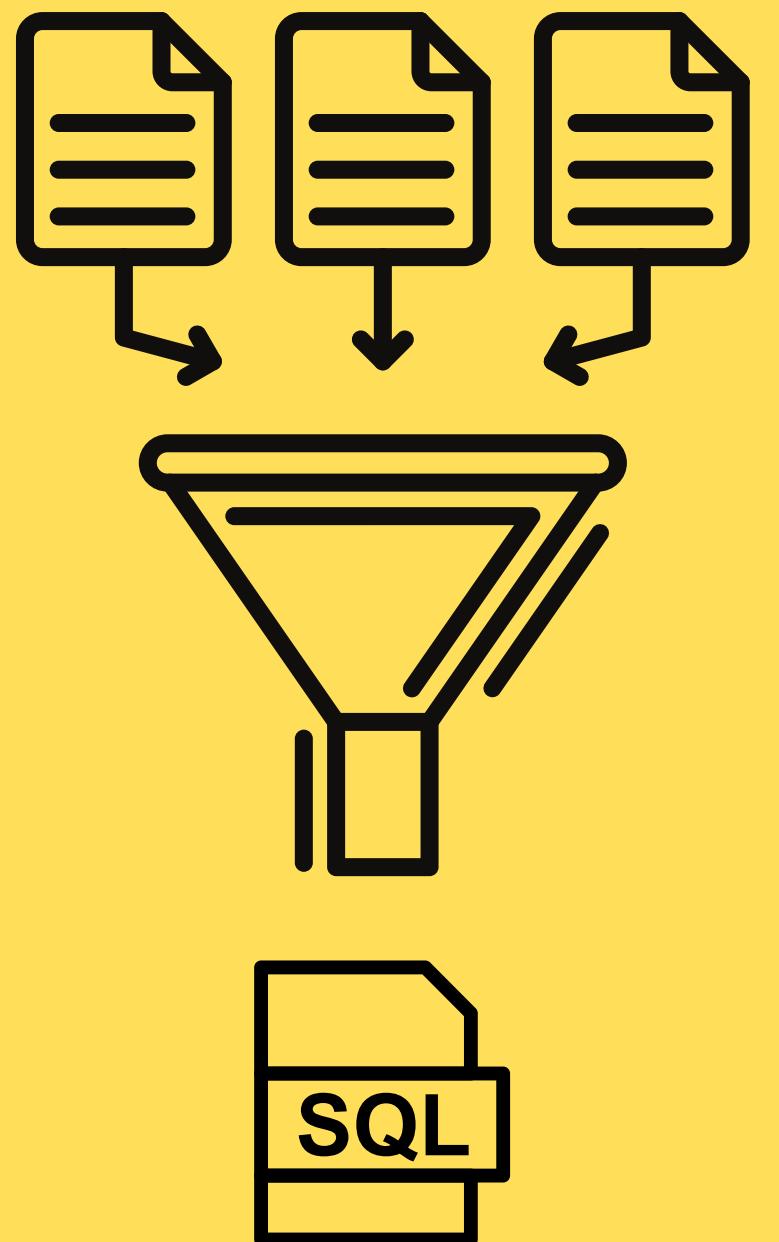


Exploring Kidney Disease

SQL Data Report



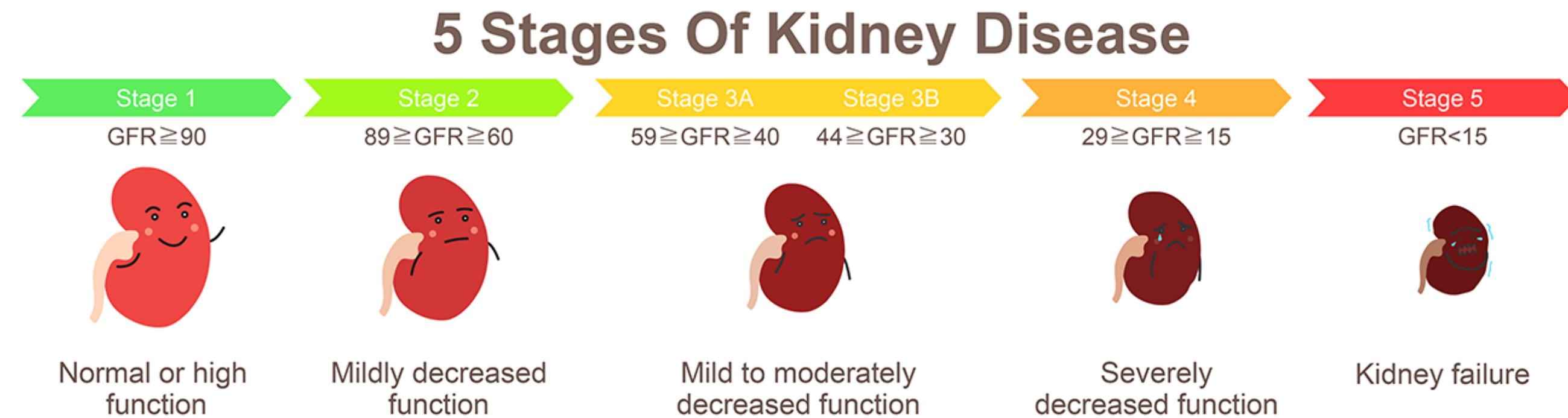
AGENDA



- 1** PREMISE
- 2** OBJECTIVE
- 3** DATA COLLECTION
- 4** DATA CLEANING
- 5** DATA ANALYSIS
- 6** INSIGHTS
- 7** ACTION ITEMS
- 8** LIMITATIONS
- 9** SUMMARY

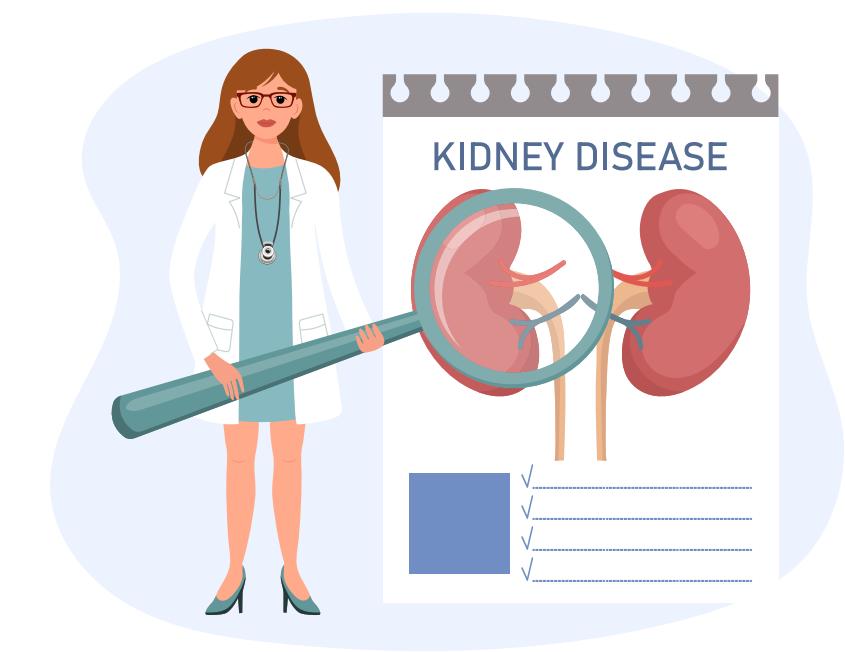
Premise

- Chronic kidney disease (CKD) is a serious medical condition that slowly deteriorates renal function
- Leads to kidney failure / dialysis if untreated
- Large proportion of the population are affected especially those with pre-existing conditions
- Demand for dialysis treatments and kidney transplants is on the rise so it is critical to understand how to prevent and manage CKD



Objective

- Write SQL queries to analyze real clinical data to calculate statistics in CKD patients
- Conduct exploratory analysis by comparing the average values of blood tests for 2 groups: CKD vs. non-CKD patients to understand the risk factors of kidney disease
- Visualize health metrics in Tableau to highlight differences between CKD and non-CKD patients



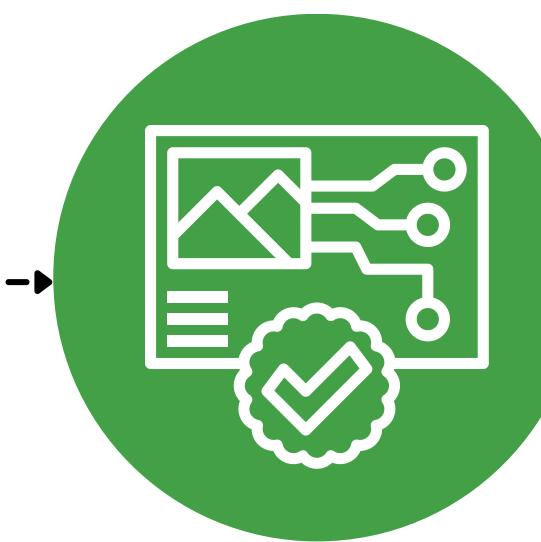
Data Collection

- The dataset was found on Kaggle with the intention of analyzing de-identified patient data containing medical attributes in order to generate health metrics for clinical research purposes



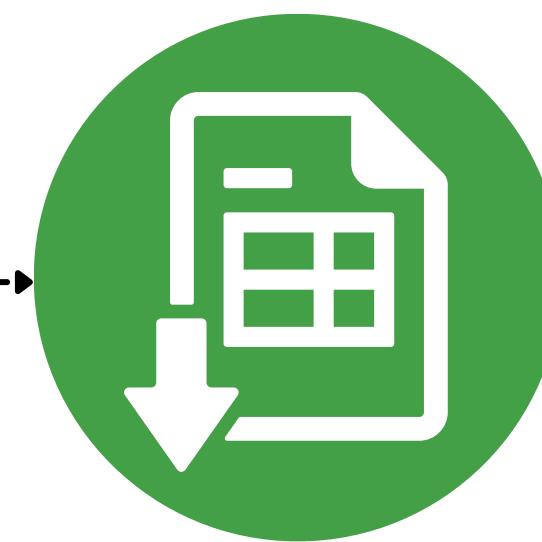
Step 1

Investigate topic of interest and goals



Step 2

Identify dataset for analysis



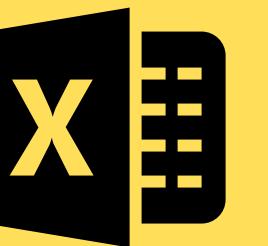
Step 3

Import file into Excel and determine data cleaning needs

Z1		X	✓	fx	classification	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W
						id	age	bp	sg	al	su	rbc	pc	pcc	ba	bgr	bu	sc	sod	pot	hemo	pcv	wc	rc	htn	dm	cad	appet
0		48	80	1.02	1	0																44	7800	5.2	yes	yes	no	good
1		7	50	1.02	4	0																38	6000	no	no	no	good	
2		62	80	1.01	2	3	normal															31	7500	no	yes	no	poor	
3		48	70	1.005	4	0	normal														32	6700	3.9	yes	no	no	poor	
4		51	80	1.01	2	0	normal														35	7300	4.6	no	no	no	good	
5		60	90	1.015	3	0															39	7800	4.4	yes	yes	no	good	
6		68	70	1.01	0	0															36		no	no	no	good		
7		24		1.015	2	4	normal														44	6900	5	no	yes	no	good	
8		52	100	1.015	3	0	normal														33	9600	4	yes	yes	no	good	
9		53	90	1.02	2	0	abnormal														29	12100	3.7	yes	yes	no	poor	
10		50	60	1.01	2	4															28		yes	yes	no	good		
11		63	70	1.01	3	0	abnormal														32	4500	3.8	yes	yes	no	poor	
12		68	70	1.015	3	1															28	12200	3.4	yes	yes	yes	poor	
13		68	70																				yes	yes	yes	poor		
14		68	80	1.01	3	2	normal														16	11000	2.6	yes	yes	yes	poor	
15		40	80	1.015	3	0															24	3800	2.8	yes	no	no	good	
16		47	70	1.015	2	0																	no	no	no	good		
17		47	80																				yes	no	no	poor		
18		60	100	1.025	0	3															37	11400	4.3	yes	yes	yes	good	
19		62	60	1.015	1	0															30	5300	3.7	yes	no	yes	good	
20		61	80	1.015	2	0	abnormal														24	9200	3.2	yes	yes	yes	poor	
21		60	90																		32	6200	3.6	yes	yes	yes	good	
22		48	80	1.025	4	0	normal														32	6900	3.4	yes	no	no	good	
23		21	70	1.01	0	0																39	8300	4.6	yes	no	no	poor
24		42	100	1.015	4	0	normal														29	8400	3.7	yes	yes	no	good	
25		61	60	1.025	0	0															35	10300	4	yes	yes	no	poor	
26		75	80	1.015	0	0															37	9600	4.1	yes	yes	yes	good	
27		69	70	1.01	3	4	normal															38		no	no	yes	good	
28		75	70			1	3																yes	yes	no	good		
29		68	70	1.005	1	0	abnormal														30	7800	4	no	no	no	poor	
30			70																		34	9600	4	yes	yes	no	poor	
31		73	90	1.015	3	0															29		yes	no	no	poor		
32		61	90	1.01	1	1															29		yes	no	no	poor		
33		60	100	1.02	2	0	abnormal														36	9800	4.9	yes	yes	no	poor	
34		70	70	1.01	1	0	normal														32		yes	no	no	good		
35		65	90	1.02	2	1	abnormal														32		yes	no	no	good		
36		76	70	1.015	1	0	normal	normal	notpresent	notpresent											32		yes	no	no	good		
38		69	80	1.02	3	0	abnormal	normal	notpresent	notpresent											32		yes	no	no	good		



Data Cleaning



File Edit View Query Project Tools Window Help Full Screen Quick Launch (Ctrl+Q)

SQLQuery1.sql - D...LAPTOP\kwjad (61)

```
***** Script for SelectTopNRows command from SSMS *****/
SELECT TOP (1000) [id]
    ,[age]
    ,[blood_pressure]
    ,[specific_gravity]
    ,[albumin]
    ,[sugar]
    ,[red_blood_cells]
    ,[pus_cell]
    ,[pus_cell_clumps]
    ,[bacteria]
    ,[blood_glucose_random]
    ,[blood_urea]
    ,[serum_creatinine]
    ,[sodium]
    ,[potassium]
    ,[hemoglobin]
    ,[packed_cell_volume]
    ,[white_blood_cell_count]
    ,[red_blood_cell_count]
    ,[hypertension]
    ,[diabetes_mellitus]
    ,[coronary_artery_disease]
    ,[appetite]
    ,[pedal_edema]
    ,[anemia]
    ,[classification]
FROM [PortfolioCKD].[dbo].[KidneyDisease]
```

100 % Results

	id	age	blood_pressure	specific_gravity	albumin	sugar	red_blood_cells	pus_cell	pus_cell_clumps	bacteria	blood_glucose_random	blood_urea	serum_creatinine	sodium	potassium	hemoglobin	packe
277	2...	48	60	1.01999998092651	0	0	normal	normal	notpresent	notpresent	112	44	1.20000004768372	142	4.90000009536743	14.5	44
278	2...	24	70	1.02499997615814	0	0	normal	normal	notpresent	notpresent	140	23	0.600000023841858	140	4.69999980926514	16.2999992370605	48
279	2...	47	80	NULL	NULL	NULL	NULL	NULL	notpresent	notpresent	93	33	0.89999976158142	144	4.5	13.3000001907349	52
280	2	55	80	1.02499997615814	0	0	normal	normal	notpresent	notpresent	130	50	1.20000004768372	147	5	15.5	41

Background tasks (Ctrl+E, Ctrl+T) | DK-LAPTOP\SQLEXPRESS (16.0 ... | DK-LAPTOP\kwjad (61) | PortfolioCKD | 00:00:00 | 398 rows

- CSV file was imported into SQL Server
- Columns were renamed for clarity to reflect full names rather than abbreviations

Data Analysis

- Count the total number of patients followed by patients with each complication grouped by their classification (CKD or notCKD)

```
...  
Total Patients Table  
  
SELECT classification,  
       COUNT(*) AS "Total Patients",  
       COUNT(CASE WHEN hypertension = 'yes' THEN 1 END) AS "Hypertension Count",  
       COUNT(CASE WHEN diabetes_mellitus = 'yes' THEN 1 END) AS "Diabetes Count",  
       COUNT(CASE WHEN coronary_artery_disease = 'yes' THEN 1 END) AS "CoronaryArteryDisease  
Count",  
       COUNT(CASE WHEN anemia = 'yes' THEN 1 END) AS "Anemia Count",  
       COUNT(CASE WHEN pedal_edema = 'yes' THEN 1 END) AS "Edema Count"  
FROM KidneyDisease  
GROUP BY classification;
```

	classification	Total Patients	Hypertension Count	Diabetes Count	CoronaryArteryDisease Count	Anemia Count	Edema Count
1	ckd	248	145	133	34	59	76
2	notckd	150	0	0	0	0	0



- We can see that non-CKD patients do not have any of the comorbidities

Data Analysis

- Calculate the percentage of CKD patients grouped by each complication

```
Percentages Table

SELECT
    COUNT(CASE WHEN hypertension = 'yes' THEN 1 END) * 100.0 / COUNT(*) AS "Percentage of
Hypertension CKD Patients",
    COUNT(CASE WHEN diabetes_mellitus = 'yes' THEN 1 END) * 100.0 / COUNT(*) AS "Percentage
of Diabetes CKD Patients",
    COUNT(CASE WHEN coronary_artery_disease = 'yes' THEN 1 END) * 100.0 / COUNT(*) AS
"Percentage CoronaryArteryDisease CKD Patients",
    COUNT(CASE WHEN anemia = 'yes' THEN 1 END) * 100.0 / COUNT(*) AS "Percentage Anemia CKD
Patients",
    COUNT(CASE WHEN pedal_edema = 'yes' THEN 1 END) * 100.0 / COUNT(*) AS "Percentage Edema
CKD Patients"
FROM KidneyDisease
WHERE classification = 'ckd';
```

	Results	Messages			
1	Percentage of Hypertension CKD Patients 58.467741935483	Percentage of Diabetes CKD Patients 53.629032258064	Percentage CoronaryArteryDisease CKD Patients 13.709677419354	Percentage Anemia CKD Patients 23.790322580645	Percentage Edema CKD Patients 30.645161290322

Data Analysis

- Count the number of patients with multiple comorbidities and group by each combination

```
Multiple Comorbidities Table

SELECT
    SUM(CASE WHEN hypertension = 'yes' AND diabetes_mellitus = 'yes' THEN 1 ELSE 0 END) AS "Hypertension & Diabetes",
    SUM(CASE WHEN hypertension = 'yes' AND coronary_artery_disease = 'yes' THEN 1 ELSE 0 END) AS "Hypertension & CoronaryArteryDisease",
    SUM(CASE WHEN diabetes_mellitus = 'yes' AND coronary_artery_disease = 'yes' THEN 1 ELSE 0 END) AS "Diabetes & CoronaryArteryDisease",
    SUM(CASE WHEN hypertension = 'yes' AND diabetes_mellitus = 'yes' AND coronary_artery_disease = 'yes' THEN 1 ELSE 0 END) AS "All: Hypertension, Diabetes, CoronaryArteryDisease"
FROM KidneyDisease;
```

	Hypertension & Diabetes	Hypertension & CoronaryArteryDisease	Diabetes & CoronaryArteryDisease	All: Hypertension, Diabetes, CoronaryArteryDisease
1	102	30	26	24

Data Analysis

- Calculate the percentages of CKD patients with each combination of multiple comorbidities

```
...  
Multiple Comorbidities Percentages  
  
SELECT  
    COUNT(CASE WHEN hypertension = 'yes' AND diabetes_mellitus = 'yes' THEN 1 END) * 100.0 /  
    COUNT(*) AS "Percentage of Hypertension & Diabetes CKD Patients",  
    COUNT(CASE WHEN hypertension = 'yes' AND coronary_artery_disease = 'yes' THEN 1 END) *  
    100.0 / COUNT(*) AS "Percentage of Hypertension & CoronaryArteryDisease CKD Patients",  
    COUNT(CASE WHEN diabetes_mellitus = 'yes' AND coronary_artery_disease = 'yes' THEN 1 END)  
    * 100.0 / COUNT(*) AS "Percentage of Diabetes & CoronaryArteryDisease CKD Patients",  
    COUNT(CASE WHEN hypertension = 'yes' AND diabetes_mellitus = 'yes' AND  
    coronary_artery_disease = 'yes' THEN 1 END) * 100.0 / COUNT(*) AS "Percentage of All  
    Comorbidities CKD Patients"  
FROM KidneyDisease  
WHERE classification = 'ckd';
```

	Results	Messages		
1	Percentage of Hypertension & Diabetes CKD Patients 41.129032258064	Percentage of Hypertension & CoronaryArteryDisease CKD Patients 12.096774193548	Percentage of Diabetes & CoronaryArteryDisease CKD Patients 10.483870967741	Percentage of All Comorbidities CKD Patients 9.677419354838

Data Analysis

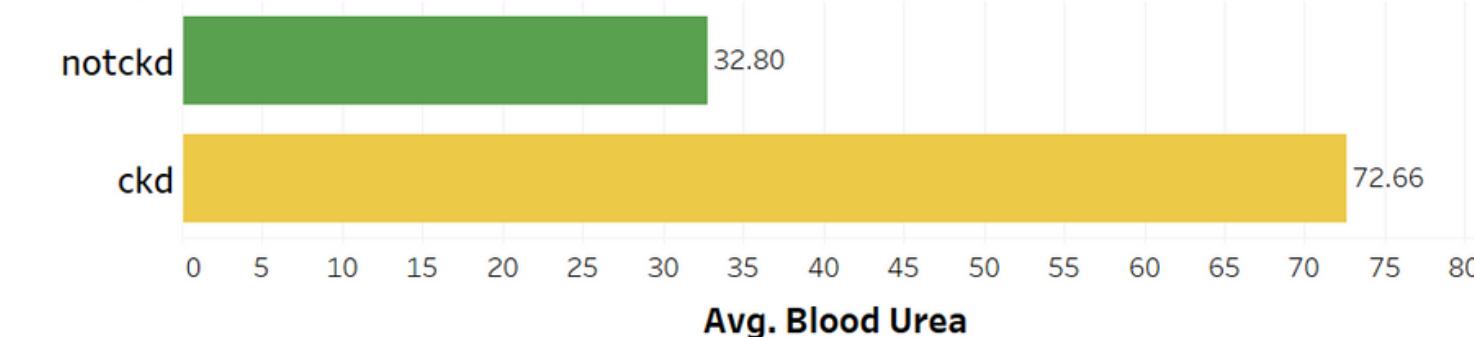
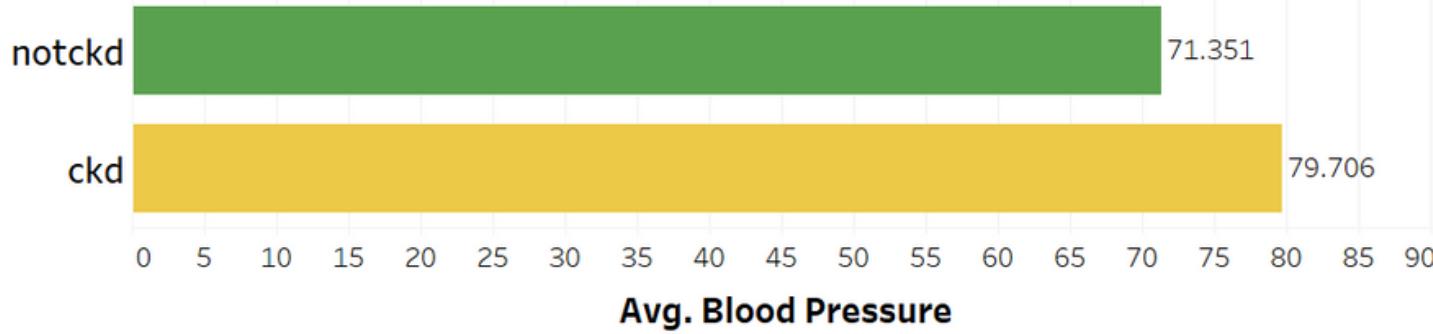
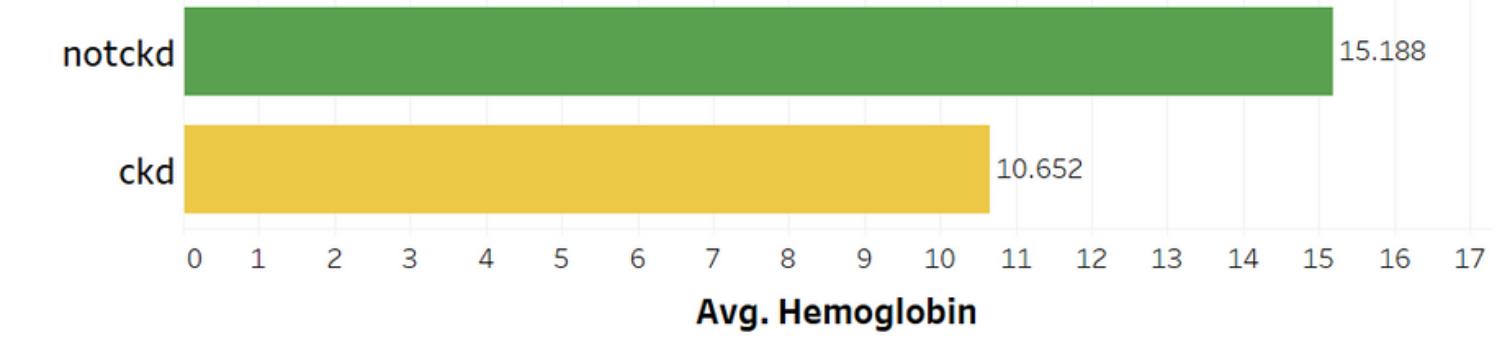
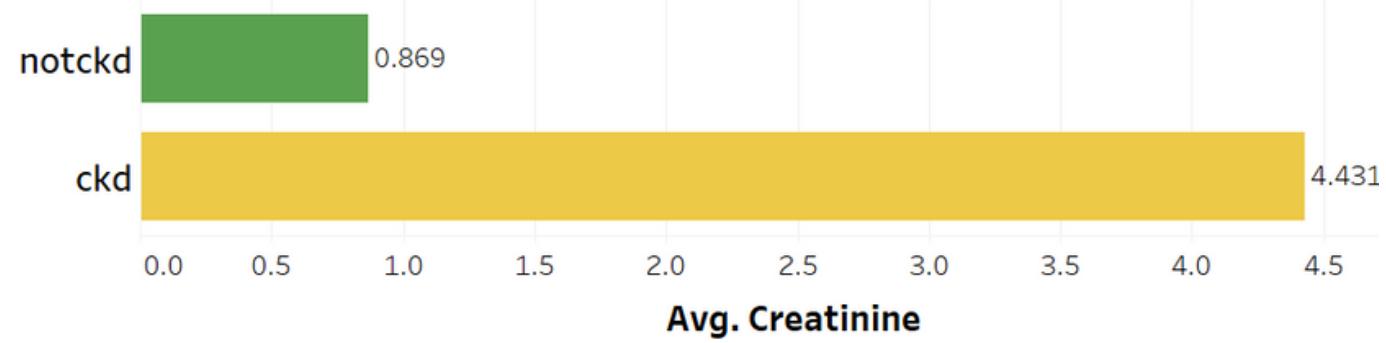
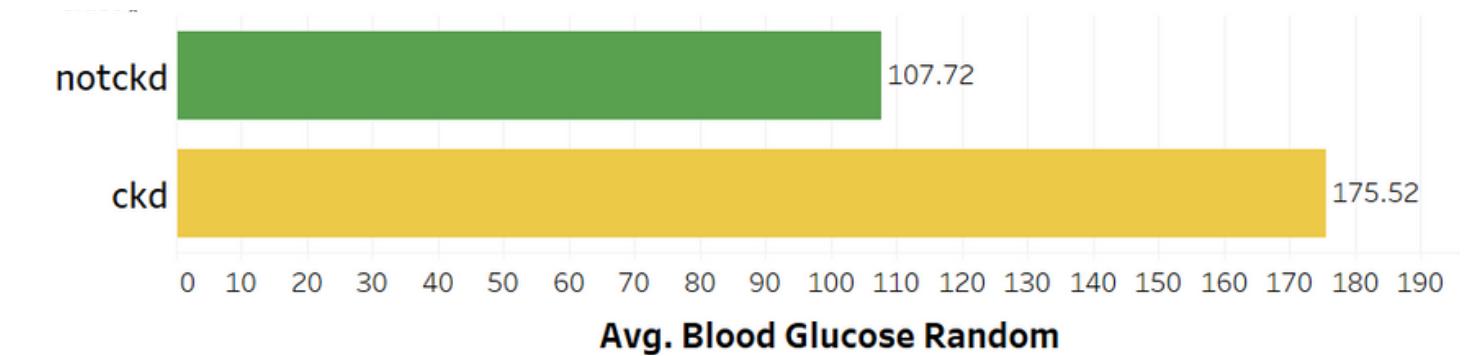
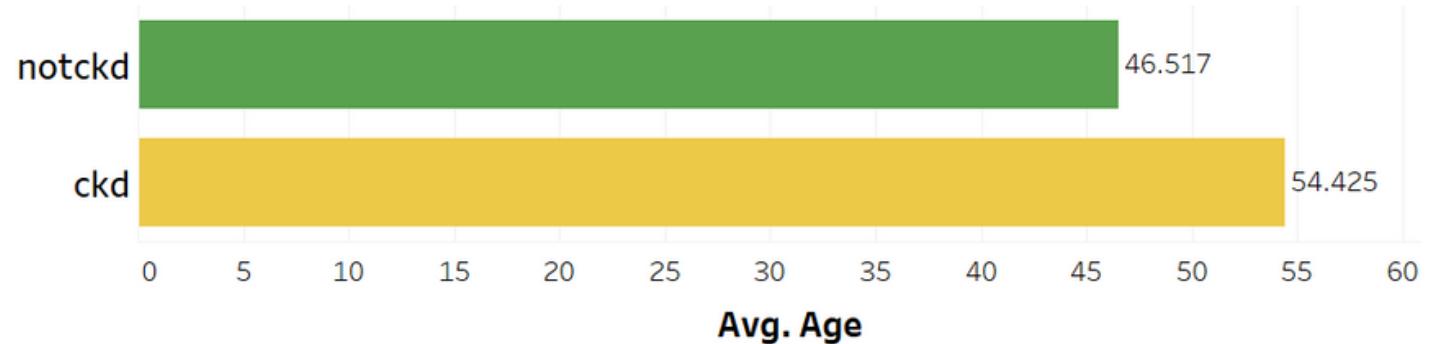
- Calculate the Averages of all relevant risk factors of CKD

```
• • • Risk Factor Averages Table

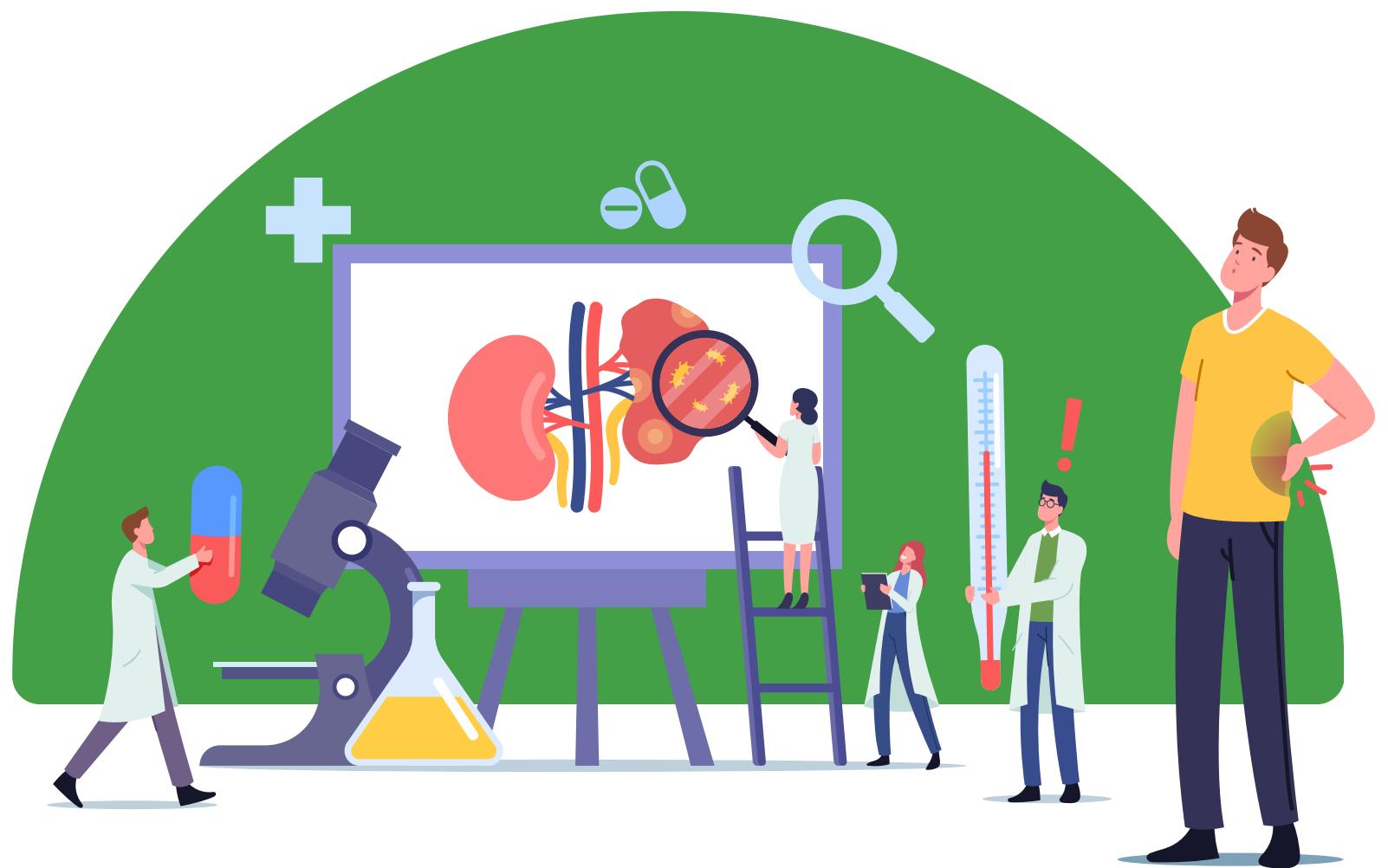
SELECT
    classification,
    AVG(age) AS "Average Age",
    AVG(serum_creatinine) AS "Average Creatinine",
    AVG(blood_pressure) AS "Average Blood Pressure",
    AVG(blood_glucose_random) AS "Average Blood Glucose Random",
    AVG(hemoglobin) AS "Average Hemoglobin",
    AVG(blood_urea) AS "Average Blood Urea",
    AVG(sodium) AS "Average Sodium",
    AVG(potassium) AS "Average Potassium"
FROM
    KidneyDisease
GROUP BY
    classification;
```

	classification	Average Age	Average Creatinine	Average Blood Pressure	Average Blood Glucose Random	Average Hemoglobin	Average Blood Urea	Average Sodium	Average Potassium
1	ckd	54	4.43072033383079	79	175	10.6522167727278	72.6561701997798	133.882530120482	4.8830303105441
2	notckd	46	0.868965529367842	71	107	15.1881944338481	32.7986111111111	141.731034482759	4.33793104434836

CKD vs. non-CKD Averages



Insights



- **Patient Demographics** - Out of 398 total patients, 248 were CKD while 150 were *non-CKD*. Average age of CKD patients was 54, while average age of *non-CKD* patients was 46
- **Creatinine** - Average creatinine level of CKD patients was 4.43 mg/dL, while average creatinine level of *non-CKD* patients was: 0.87 mg/dL
- **Hypertension** - Average blood pressure of CKD patients was 79, while average blood pressure of *non-CKD* patients was 71
 - 58.5% of CKD patients had hypertension
- **Diabetes** - Average blood glucose level of CKD patients was 175, while average Blood Glucose level of *non-CKD* patients was 107
 - 53.6% of CKD patients had diabetes
- **Anemia** - Average hemoglobin level of CKD patients was 10.6 g/dL, while average hemoglobin level of *non-CKD* patients was 15.2
 - 23.8% of CKD patients had anemia

Action Items

Healthcare Providers



- **Monitor Kidney Function:** for early detection and management of risk factors for CKD in patients, especially in those with hypertension, diabetes, and cardiovascular disease
- **Order Routine Blood Tests:** and check blood pressure for patients to evaluate medical attributes associated with kidney function including creatinine, high blood pressure, hemoglobin, sodium & potassium levels
- **Encourage Lifestyle Changes:** including maintaining a healthy weight and diet as well as engaging in regular physical activity to reduce the risk of CKD

Action Items

Researchers



- **Conduct Clinical Trials:** for further research to understand the complex interconnected relationships between age, hypertension, diabetes, and cardiovascular disease and their contribution to the progression of CKD
- **Investigate Genetics:** and the potential role of racial and genetic differences in the development and progression of CKD including studying unique characteristics of CKD in different racial groups and ensuring medical equality in care for all patients
- **Improve Data Collection / Analysis:** Clinical Data Analysts can work on standardizing protocols for data collection, using advanced statistical methods, and using electronic health records to track patient outcomes over time

Action Items



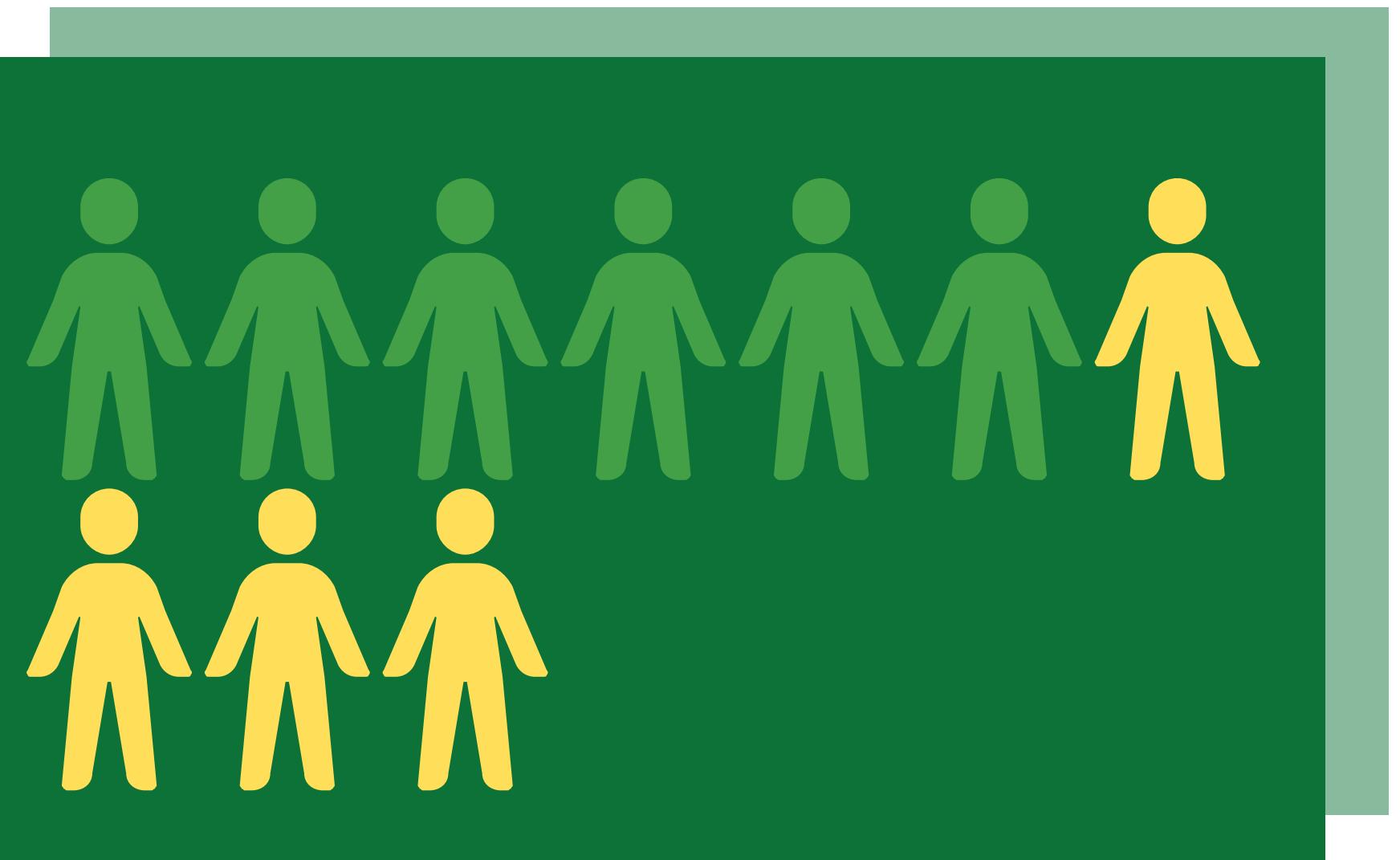
Policy Makers

- Promote Innovative Payment Models: which incentivize quality care and improve patient outcomes, such as value-based payment models
- Address Social Determinants of Health: that contribute to CKD disparities, such as poverty and lack of access to healthcare in addition to developing plans that ensure equitable access to quality care for all CKD patients, regardless of race, ethnicity, or socioeconomic status
- Launch Awareness Campaigns: to educate high-risk populations of the factors that contribute to CKD and how they can take practical steps to prevent its development

Limitations

Single Study

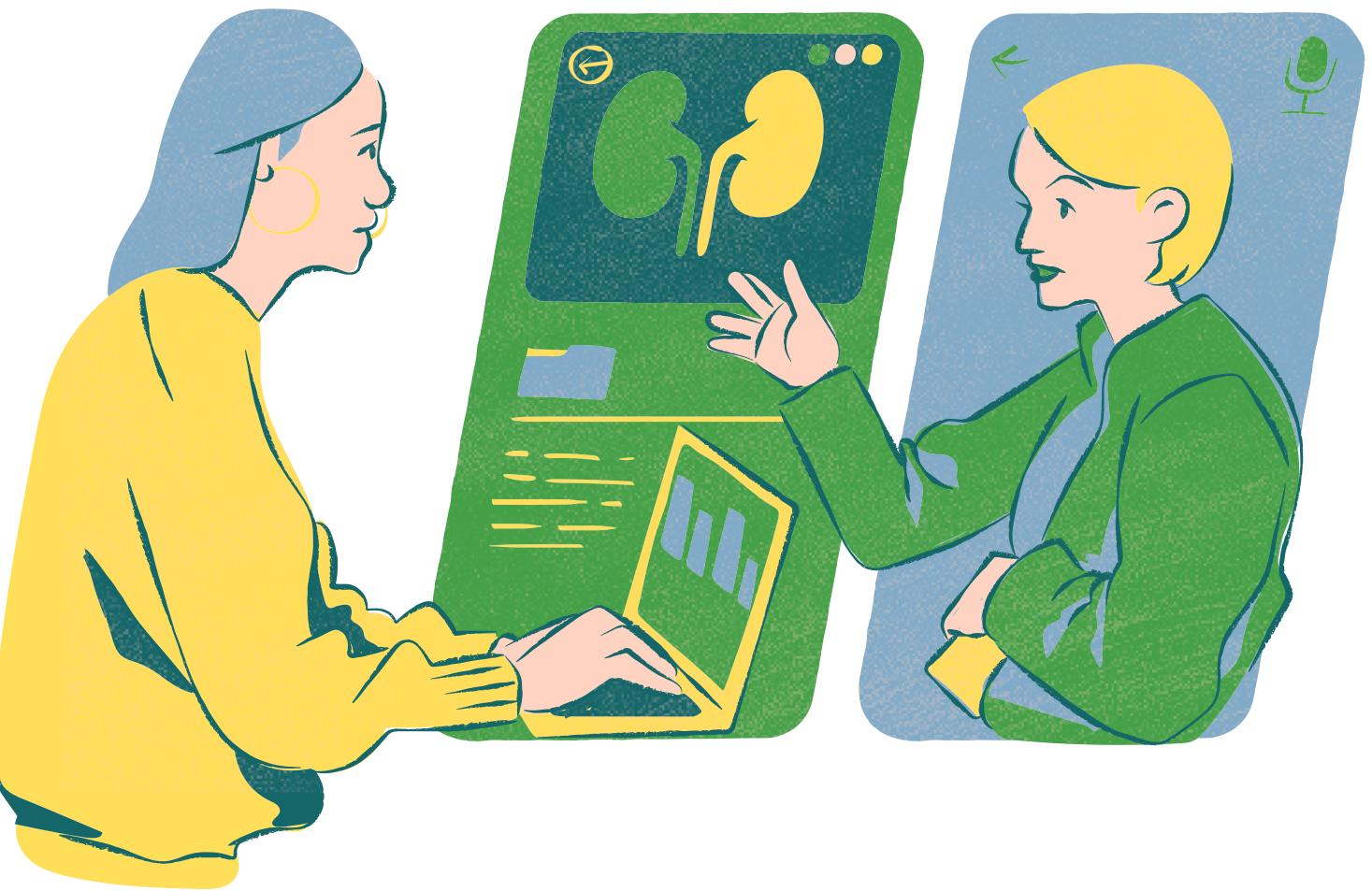
- This analysis is based on a single dataset, which does not represent the entire population of CKD patients
- Genetic, socioeconomic, and regional differences impact CKD progression & treatment
- Conducting a meta-analysis that combines multiple datasets from diverse sources would provide a more clear picture of CKD factors



Limitations

Lack of CKD Stage Specification

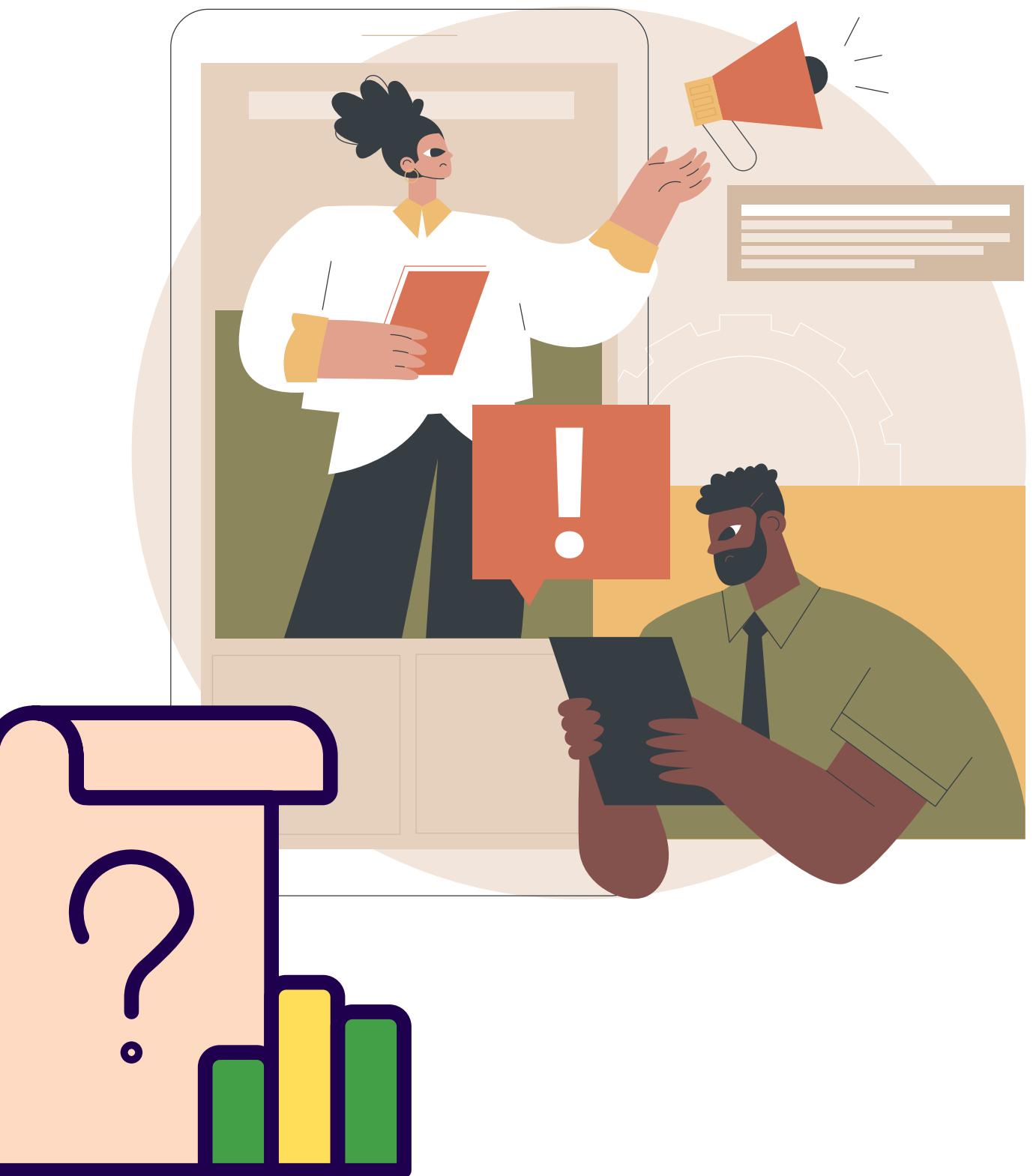
- This dataset does not specify the stage of CKD for each patient. CKD is a progressive disease that encompasses stages 1 to 5
- The absence of stage-specific information limits our understanding of how the risk factors may vary across different stages of CKD
- Analyzing data specific to each CKD stage would provide a nuanced understanding of disease progression



Limitations

Potential Confounding Variables

- The dataset used for this analysis may have inherent confounding variables that were not accounted for
- Factors like gender, socioeconomic status, family history, diet and lifestyle habits can influence both the development of CKD and the presence of comorbidities





Summary



- Overall, this data analysis highlights the significant differences in blood tests for CKD vs. non-CKD patients and the presence of comorbidities on the development of kidney disease
- CKD patients tend to be older - need for regular monitoring of kidney function for elderly patients especially those with established comorbidities
- Higher levels of creatinine, blood pressure, glucose, & uric acid - indicates decreased kidney function
- Lower levels of hemoglobin, packed cell volume, & red blood cell count - indicates a high prevalence of anemia
- Hypertension, diabetes, and coronary artery disease were the most prevalent comorbidities while edema and anemia were frequent complications of CKD
- Through concerted efforts from physicians, researchers, and policy makers, we can work together to improve the lives of those affected by kidney disease

THANK YOU

Danial Khawaja

kwja.danial360@gmail.com

daniel-khawaja.github.io/DanielKhawaja.github.io

