USING DATA ANALYTICS IN HEALTHCARE

Helping CareMore Hospital Increase Profits by Improving Diabetic Patient Outcomes

Agenda

- Business Understanding
- Predictive Analytics
- Descriptive Analytics
- Recommendation
- Cost & Benefit Analysis

CareMore Hospital can increase profits by identifying high-risk patients and providing preventative care

Current Problem

CareMore hospital is looking for ways to increase their revenues. Specifically, there is a business opportunity in improving their care for diabetic patients.

Key Objectives

Increase future profit by improving patient outcomes and satisfaction through preventative measures

Save money on potential penalties associated with high readmissions

Business Understanding

Utilize **predictive analytics** to identify high-risk diabetic patients.

Utilize **descriptive analytics** and identify key patient personas

Utilize **prescriptive analytics** to create recommendations based on personas identified above

Our analytical efforts will help CareMore better serve their diabetic patients and benefit from increased revenues

Random Forest Model

Obtained the highest predictive accuracy of **64.37%**

Clustering Analysis

Identified **4 patient personas** with specific remediation recommendations

Net Cost-Benefit Analysis

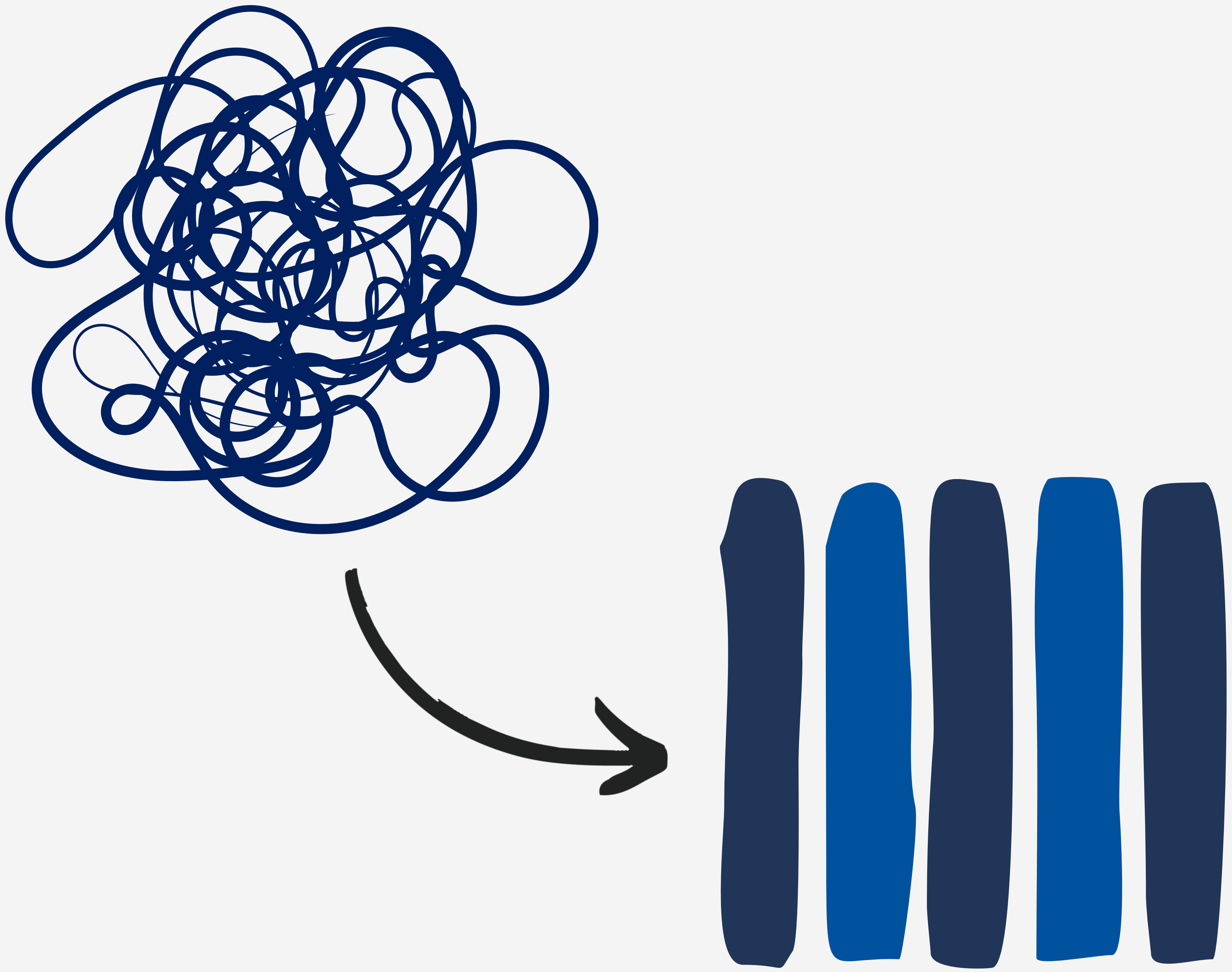
Quantified a **net benefit of $5,321,123** by leveraging predictive & prescriptive model solution

Limiting the modeling data to the most important attributes helped obtain more valuable results with higher accuracy

Dummy-coding categorical variables with several outcomes (700+ categories)

- Diagnoses Variables (ex. diag_1)
- Race (ex. Other)

Using **Weight by Information Gain** to limit modeling data to 100 top attributes



Most Important Attributes Examples:

- Insulin = No
- Number of emergency
- Number of diagnoses
- Number of outpatient
- Diag_1 = 428
- Number of medications

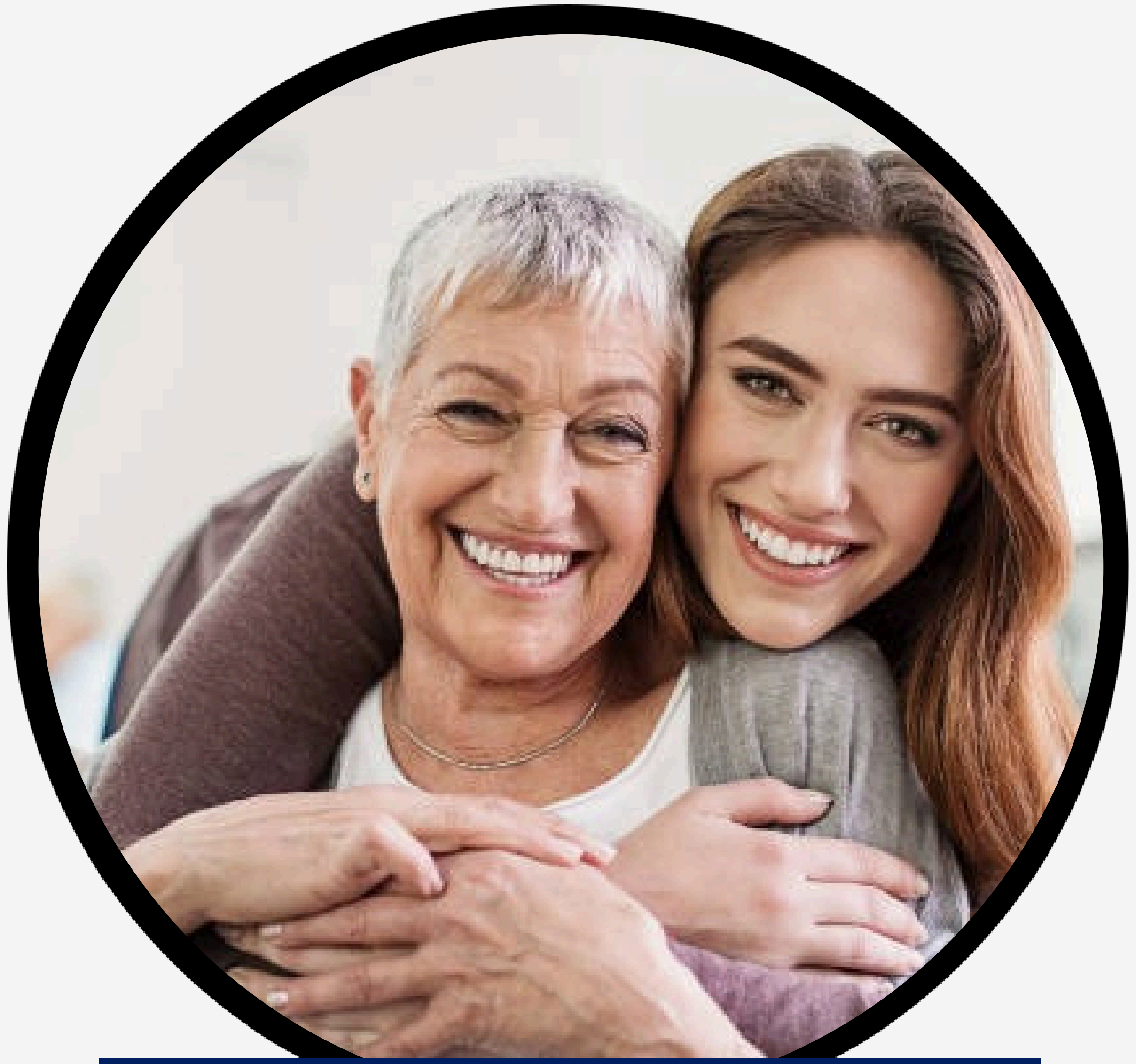
A Random Forests model yielded the highest testing accuracy of 64.37% in predicting diabetic patient readmission

Why Random Forests Model?

Model	Precision	Recall	F-1 Measure	Testing Accuracy
Decision Tree Model	65.40%	43.56%	52.30%	62.87%
Bagging Model	64.03%	52.77%	57.85%	63.87%
Random Forests Model	65.63%	50.78%	57.30	64.37%

Readmitted patients can be clustered into four personas based on race, age, gender, and diagnoses

Meet Grandma Sally who represents the most common high-risk patient and is susceptible to having chronic bronchitis



Grandma Sally

41.64%

of Readmitted Patients

Patient Type

Patients like **Grandma Sally** are elderly, *Caucasian women* that are most likely to develop heart failure due to a decline in estrogen, which leads to more bad cholesterol

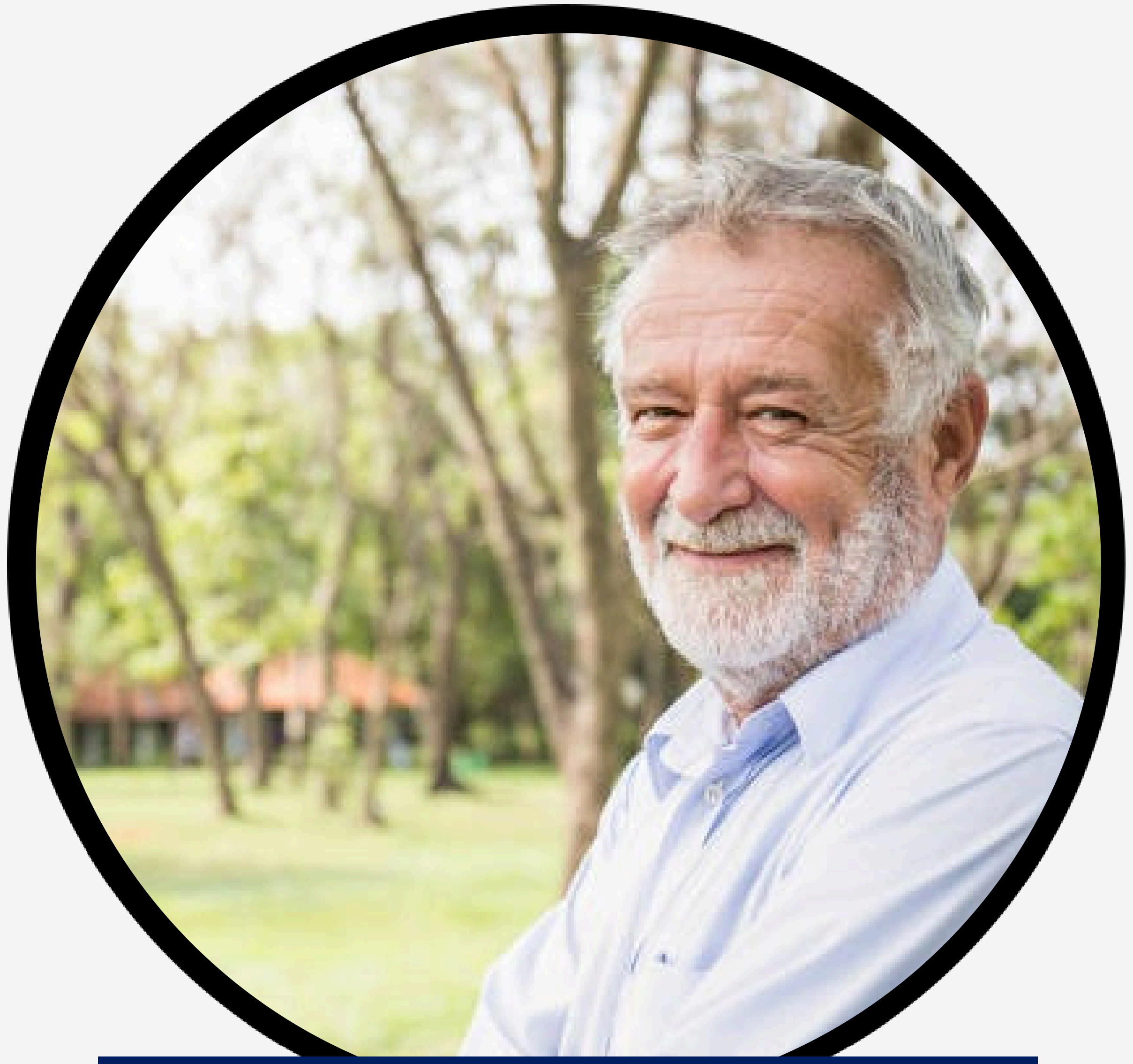
Diabetes relates to bronchitis because of **chronic inflammation** due to **damaging blood vessels**

Respiratory illnesses also may be risk factors for diabetes

Diagnosis Type

Grandma Sally is most likely to have **ischemic heart disease, heart failure,** and **chronic bronchitis**

Meet Uncle Jerry who represents the second most common high-risk patient prone to having surgical complications



Uncle Jerry

36.48%

of Readmitted Patients

Patient Type

Patients like **Uncle Jerry** are elderly, *Caucasian men* who are likely to have several health troubles related to diabetes and twice as likely for having a heart disease or stroke

---

Abnormally high levels of LDL cholesterol contribute to heart attacks

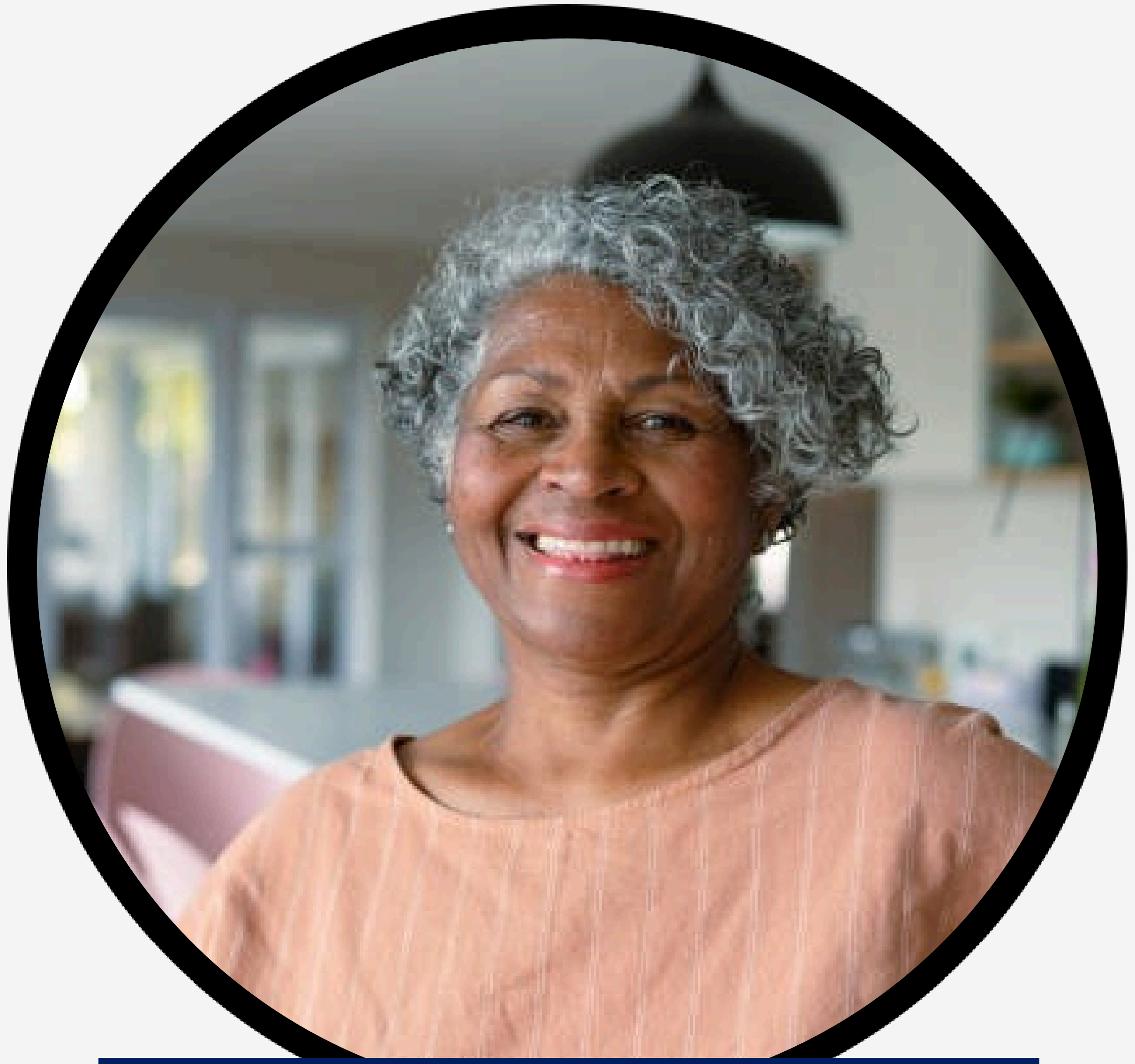
---

High glucose levels from diabetes **damage blood vessels** that control your heart

Diagnosis Type

Uncle Jerry is most likely to have **heart attacks**, **surgical complications**, **ischemic disease**, and **heart failure**

Meet Aunt Anne who represents the third common high risk patient prone to heart disease and heart failure



Aunt Anne

13.01%

of Readmitted Patients

Patient Type

Patients like **Aunt Anne** are typically *middle-aged to elderly*, *African American* women.

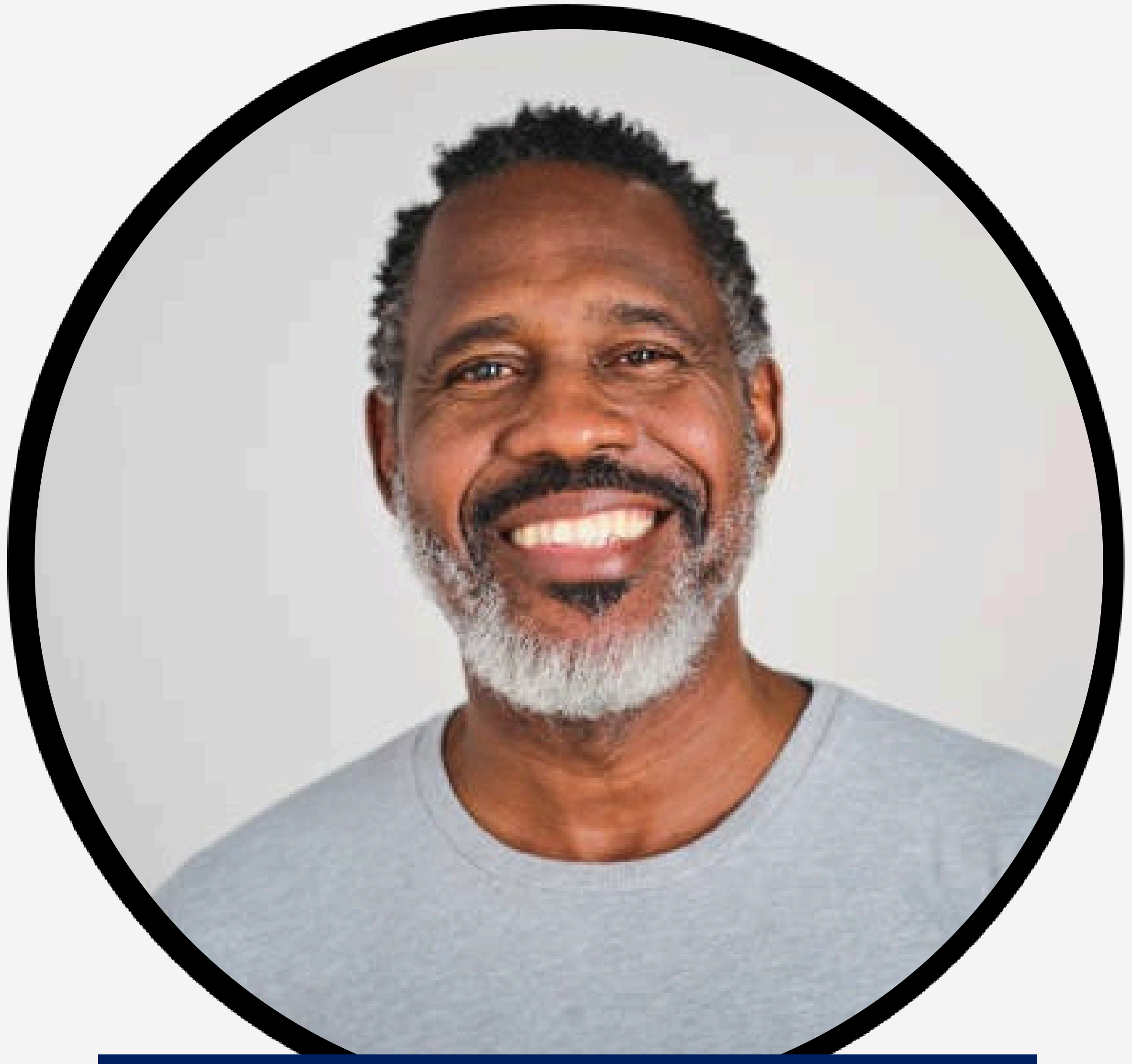
Diabetes in women **increase the risk of heart disease** than it does in **men**, due to added risk factors like obesity

High glucose levels from diabetes **damage blood vessels** that control your heart, as well as pulmonary vessels

Diagnoses Type

Aunt Anne is most likely to have **respiratory symptoms** and **heart disease** and **failure**.

Meet Mr. James who represents least common high-risk patient and is likely to have pancreatitis and heart problems



Mr. James

8.08%

of Readmitted Patients

Patient Type

Patients like **Mr. James** tend to be *middle-aged to elderly*, *African American* men with previous medical complications, and twice as likely for **heart disease** or stroke

---

Pancreatitis is correlated with diabetes because of an inability to properly absorb insulin

---

Men are more likely to be **predisposed** to **heart attacks**

Diagnosis Type

Mr. James is most likely to have **pancreatitis** and **heart failure**

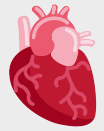
Implement customized proactive care to each identified cluster

1. Identify high-risk patients with Random Forests 2. Cluster high risk patients 3. Provide customized proactive care to each cluster

Aunt Anne

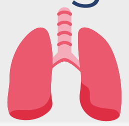
Heart failure

- RN patient education
- Support at discharge
- Improve communication
- Early follow-up [1]



Respiratory Symptoms

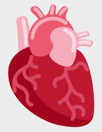
- Patient education [2,3]
- Discharge care bundles w/ personalized patient action plan [3]
- Recommend pulmonary rehab programs [2]



Grandma Sally

Heart problems

- Support at discharge
- Improved communication
- Early follow-up [1]



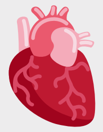
Bronchitis

- Informational pamphlet on recommended discharge care [4]

Uncle Jerry

Heart Failure

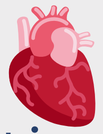
- RN patient education
- Support at discharge
- Improved communication
- Early follow-up [1]



Mr. James

Heart Failure

- RN patient education
- Support at discharge
- Improved communication
- Early follow-up [1]



Pancreatitis

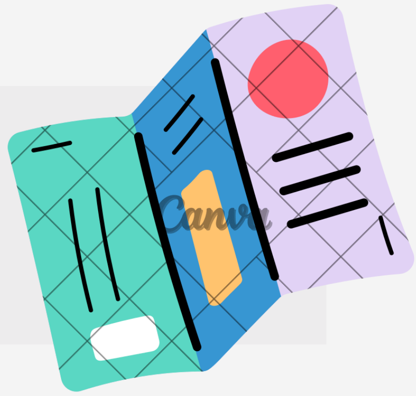
- Scan for cholecystectomy
- Alcohol cessation interventions [5]

Diabetes: Inpatient Diabetic Management Service (IDMS) - endocrinologist, nurse practitioner & diabetes educator provide patient education and action plan [6]

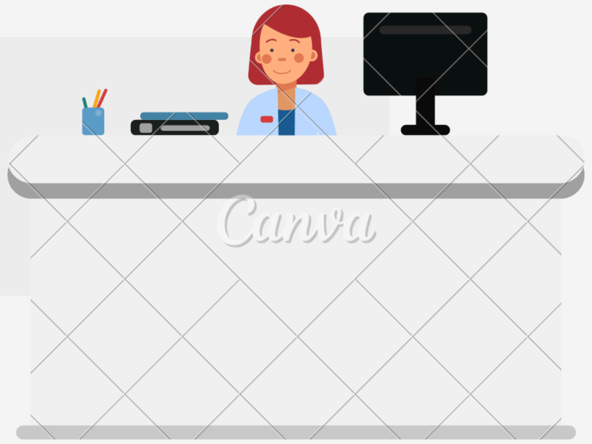
Next steps include informing relevant doctors, creating educational pamphlets, and training desk staff

1. Inform RNs, Endocrinologists, Nurse Practitioners & Diabetes Educators of their roles in new proactive care and and gain their feedback

2. Purchase existing or develop new diabetic educational pamphlets



3. Train desk staff on improved communication, increased discharge support and scheduling early follow ups



A net cost-benefit matrix shows a potential savings up to $7,451 and cost up to $327 per patient

- Reduction of readmission fee: **-$7,600**

- Increased future revenue from improved customer satisfaction: **-$330**

- Cost of customized proactive care: **$327**

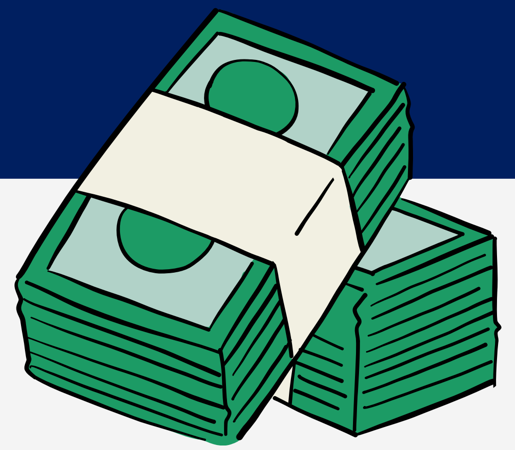
- Cost of customized proactive care: **$327**

	Predicted class	
Actual class	**Readmitted**	**Not readmitted**
Readmitted	**-$7,603**	**$0**
Not readmitted	**$327**	**$0**

Our recommendation has a net benefit of $5,321,123 resulting from confusion matrix combined with the cost matrix

		Predicted class	
Actual class		Readmitted	Not readmitted
	Readmitted	716	694
	Not readmitted	375	1,215

$5,321,123
Net benefit



Based on cluster analysis and random forest modeling, CareMore hospital has potential to increase revenues

01 Random Forest Modeling

Use random forest model to predict and identify patient likelihood to be readmitted

02 Cluster Analysis

Based on readmission outcome, use cluster analysis to classify into one of four personas

03 Identify Persona & Care

Based on persona, use relevant prescription to aid in providing adequate care

04 Net Benefit

Resulting improvement in patient care and satisfaction has an expected net benefit of over $5.3 million dollars.

THANK YOU!

Q&A

COMM 4522
Team 2
Nora Bame, Yani Iben, Elijah McFarland,
Ashley Pham, Hannah Ventura
12/4/2024

Appendix Directory

Overview

- CareMore Hospital can increase profits by identifying high-risk patients and providing preventative care
- Our analytical efforts will help CareMore better serve their diabetic patients and benefit from increased revenues

Predictive Analytics

- Limiting the modeling data to the most important attributes helped obtain more valuable results with higher accuracy
- A Random Forests model yielded the highest testing accuracy of 64.37% in predicting diabetic patient readmission

Descriptive Analytics

- The cluster analysis is supported by medical context
- Readmitted patients can be clustered into four personas based on race, age, gender, and diagnoses
- Grandma Sally, Uncle Jerry, Aunt Anne, & Mr. James

Recommendation

- Implement customized proactive care to each identified cluster
- Next steps include informing relevant doctors, creating educational pamphlets, and training desk staff

Evaluation

- A net cost-benefit matrix shows a potential savings up to $7,451 and cost up to $327 per patient
- Our recommendation has a net benefit of $5,321,123 resulting from confusion matrix combined with the cost matrix

Appendices

Appendix A: Cost Matrix - cost & benefit breakdown

Reduction of readmission fee: **-$7,600**

- $15,200 cost of readmission per patient on average [7]
- 5-79% of readmissions are deemed avoidable [8]
- 31-69% reduction in risk of readmission from IDMS Co-management [9]
- education and medication review reduced heart failure readmissions by 56.2% [10]
- *estimate 50% reduction in readmission**
- *$15,200 * 50% = $7,600*

**Italics indicate our assumptions/calculations*

Appendix A: Cost Matrix - cost & benefit breakdown

Increased future revenue from improved customer satisfaction: **-$330**

- *increase in customer satisfaction for the **50%** readmissions prevented**
- *increased customer satisfaction increases customer retention rate*
- increasing customer retention rates by 5% increases profits by 25% to 95% [11]
- *estimate **40%** increase in profit for non-readmission*
- National average ~ **$15,000** revenue per admission [12]
- average profit margin **11%** [13]
- *50%*40%*$15,000*11%= $ 330*

**Italics indicate our assumptions/calculations*

Appendix A: Cost Matrix - cost & benefit breakdown

Cost of customized proactive care: **$327**

<u>All clusters:</u> Diabetic preventative care: **$270**

- IDMS [6]: $120 Endocrinologist [14], $100 NP [15], $50 Diabetic Educator [16]

<u>Cluster 1 care:</u> $120 (13%)

- Pulmonologist's time to make discharge care bundle: $100 [17]
- 30 min RN time for patient education: $20 [18]

<u>Cluster 2 care:</u> $16 (42%)

- educational pamphlet: $1 [19]
- 1 hour hospital front desk staff for improved communication and support at discharge: $15 [20]

<u>Cluster 3 care:</u> $35 (36%)

- 30 min of RN time for patient education ($20) [18]
- 1 hour hospital front desk staff for improved communication and support at discharge ($15) [20]

<u>Cluster 4 care:</u> $247 (9%)

- 30 min RN time for education: $20 [18]
- 1 hr hospital desk staff for improved communication & support at discharge: $15 [20]
- abdominal ultrasound for Cholecystectomy: $205 [21,22]
- 30 min RN time [18] for alcohol cessation intervention for 1/3 of patients w/ alcohol-induced pancreatitis ($20/3 = $7) [23]

*270+($120*13%)+($16*42%)+($35*36%)+($247*9%)=$327*

Appendix B: Decision Tree Model Trials

Model	Parameters	Training Accura	Testing Accura	Notes
Decision Tree	Maximal depth = 80, minimal leaf size = 1, m	65.49%	62.87%	Used optimize parameters w/regularization & lambda rar
Decision Tree	criterion = information_gain; maximal depth =	-	60.83%	New Dataset
Decision Tree	DT(criterion = info_gain, maximal depth = 10	66.34%	60.83%	New Dataset
Decision Tree	DT(criterion = info_gain, maximal depth = 10	66.34%	60.83%	New Dataset
Decision Tree	DT(criterion = info_gain, maximal depth = 10	62.71%	60.60%	New Dataset
Optimized Decis	Decision Tree.maximal_depth = 0 Decision Tree.confidence = 0.25 Decision Tree.minimal_size_for_split = 1 Decision Tree.minimal_leaf_size = 31 Decision Tree.apply_pruning = true		60.57%	New Dataset
Decision Tree	Maximal depth = 2, minimal gain = .0001, minimal leaf size =		60.03%	New Dataset
Decision Tree	Maximal depth = 33, minimal gain = .1, minimal leaf size = 2,		60.03%	New Dataset
Decision Tree	Maximal depth = 38, minimal gain = .08, minimal leaf size = 2,		60.02%	New Dataset
Decision Tree	Maximal depth = 5, minimal gain = .0001, minimal leaf size =		60.02%	70/30 split new data
Decision Tree	DT(criterion = info_gain, maximal depth = 25	85.81%	55.93%	Does not have encounter_id and patient_nbr
Decision Tree	DT(criterion = info_gain, maximal depth = 50	91.41%	55.90%	Does not have encounter_id and patient_nbr
Decision Tree	maximal_depth= 18, confidence= 0.06699999999999999, mir		53.63%	
Decision Tree	Maximal depth = 10, minimal gain = .001, mi	53.60%	53.43%	Overfitting
Decision Tree	Maximal depth = 8, minimal gain = .0005, mi	53.60%	53.43%	Overfitting
Decision Tree	Maximal depth = 10, minimal leaf size = 2, m	53.54%	53.30%	

Appendix C: Random Forest Model Trials

Model	Parameters	Training Accura	Testing Accura	Notes
Random Forest	number_of_trees =100, criterion = informatio	-	64.37%	New Dataset - optimized
Random Forest	number_of_trees =100, criterion = informatio	-	64.27%	New Dataset
Random Forest	number_of_trees =100, criterion = informatio	-	63.77%	New Dataset
Random Forest	number_of_trees =100, criterion = informatio	-	63.67%	New Dataset optimized
Random Forest	Base classifer: DT(criterion = info_gain, maximal depth = 15,		63.67%	New Dataset optimized
Random Forest	Base classifer: DT(criterion = info_gain, maximal depth = 20,		62.67%	Used optimize parameters
Random Forest	Base classifer: DT(criterion = info_gain, maximal depth = 30,		62.50%	New Dataset
Random Forest	Base classifer: DT(criterion = info_gain, maximal depth = 30,		62.47%	New Dataset
Random Forest	Base classifer: DT(criterion = info_gain, maximal depth = 15,		62.40%	New Dataset
Random Forest	number_of_trees =100, criterion = informatio	-	62.33%	New Dataset
Random Forest	Base classifer: DT(criterion = info_gain, maximal depth = 25,		62.33%	New Dataset
Random Forest	number_of_trees =100, criterion = informatio	-	62.30%	New Dataset
Random Forest	Base classifer: DT(criterion = info_gain, maximal depth = 20,		62.27%	New Dataset
Random Forest	Base classifer: DT(criterion = info_gain, max	-	62.17%	Took 7 minutes to process
Random Forest	number_of_trees =100, criterion = informatio	-	61.87%	New Dataset
Random Forest	Base classifer: DT(criterion = info_gain, max	-	53.00%	
Random Forest	Base classifer: DT(criterion = info_gain, maximal depth = 30,		53.00%	

Appendix D: Ensemble Models Trials

Model	Parameters	Training Accura	Testing Accura
Vote Ensemble	Base classfiers = 2; DT(criterion = info_gain,	-	61.73%
Vote Ensemble	Base classfiers = 3; DT(criterion = info_gain,	-	62.30%
Vote Ensemble	Base classfiers = 3; DT(criterion = info_gain,	-	61.77%
Vote Ensemble	Base classfiers = 3; DT(criterion = info_gain,	-	61.40%
Vote Ensemble	Base classfiers = 3; DT(criterion = info_gain,	-	61.40%
Vote Ensemble	Base classfiers = 2; DT(criterion = info_gain,	-	62.17%
Vote Ensemble	Base classfiers = 2; DT(criterion = info_gain,	-	61.73%

Model	Parameters	Training Accura	Testing Accura
Bagging	(sample_ratio = 0.5, iterations = 100), DT (cr	-	63.87%
Bagging	(sample_ratio = 0.5, iterations = 100), DT (cr	-	63.70%
Bagging	(sample_ratio = 0.5, iterations = 100), DT (cr	-	63.60%
Bagging	(sample_ratio = 0.8, iterations = 100), DT (cr	-	63.20%
Bagging	(sample_ratio = 0.8, iterations = 100), DT (cr	-	63.13%
Bagging	(sample_ratio = 0.8, iterations = 100), DT (cr	-	60.23%
GradientBoostin	number_of_trees = 90, maximal_depth = 10,	-	62.72%
GradientBoostin	number_of_trees = 100, maximal_depth = 1(	-	62.50%
AdaBoost	iterations = 10, DT(criterion = information_ga	-	61.50%
AdaBoost	iterations = 8, DT(criterion = information_gai	-	61.30%
GradientBoostin	number_of_trees = 100, maximal_depth = 5,	-	61.27%
AdaBoost	iterations = 10, DT(criterion = information_ga	-	61.23%
AdaBoost	iterations = 10, DT(criterion = information_ga	-	60.60%

Appendix E: KNN & Logistic Regression Model Trials

Model	Parameters	Training Accura	Testing Accura
KNN	Weighted vote = yes; measure type = numer	68.53%	56.17%
KNN	Weighted vote = yes; measure type = numer	62.14%	56.03%
KNN	Weighted vote = yes; measure type = numer	62.11%	56.00%
KNN	Weighted vote = yes; measure type = numer	64.03%	55.83%
KNN	Weighted vote = yes; measure type = numer	63.10%	55.60%
KNN	Weighted vote = yes; measure type = numer	65.36%	55.10%
KNN	Weighted vote = yes; measure type = numer	65.34%	55%
KNN	Weighted vote = yes; measure type = numer	71.14%	54.73%

Model	Parameters	Training Accura	Testing Accura
Logistic Regress	Solver = auto, Use Regularization = true, lambda = 0.0152		58.70%
Logistic Regress	Solver = auto, Use Regularization = true, lambda = 0.0152		58.57%
Logistic Regress	Solver = auto, Use Regularization = true, lambda = 0.0152		58.57%
Logistic Regress	Solver = auto, Use Regularization = true, lambda = 0.016		58.27%
Logistic Regress	Solver = auto, Use Regularization = true, lambda = 0.100		58.13%

Appendix F: Data Preprocessing for Cluster Analysis

Removed Identifiers

Excluded information like patient number and ID that were identifiers

Weighted Attribues by Information Gain

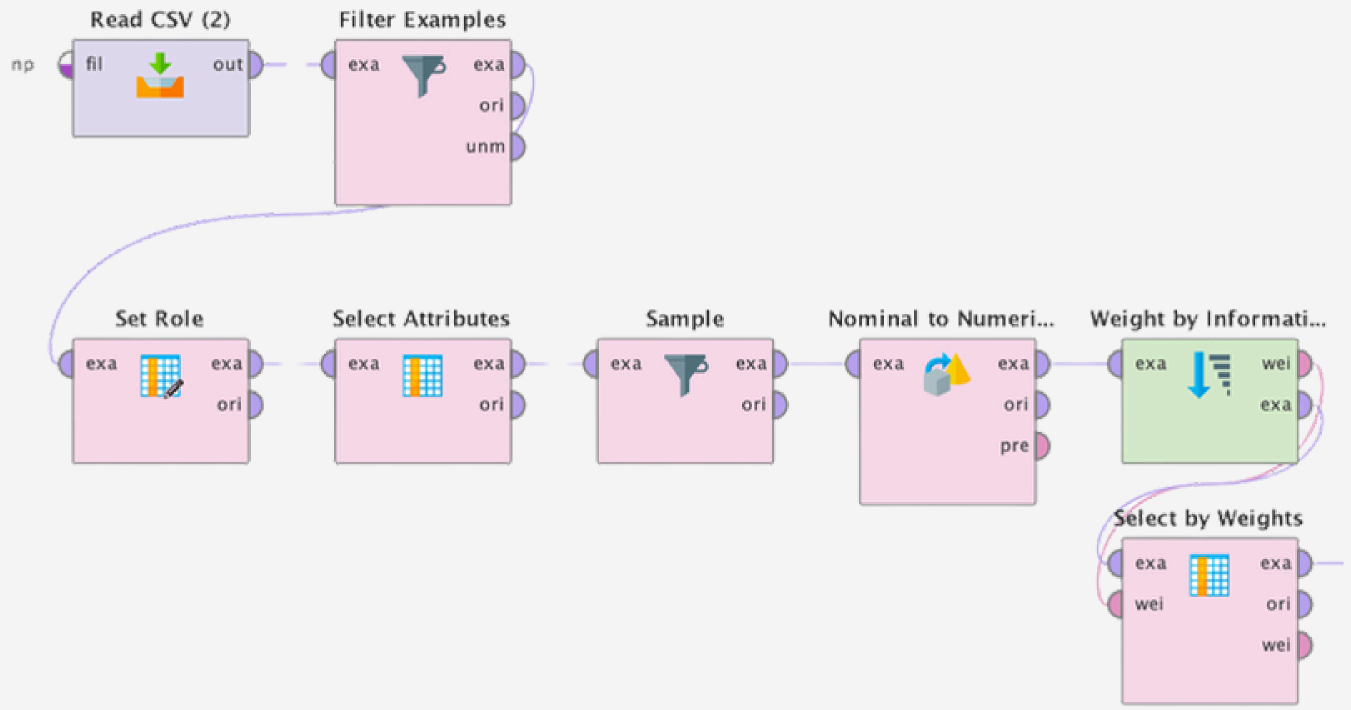
Only included attributes leading to the most information gain

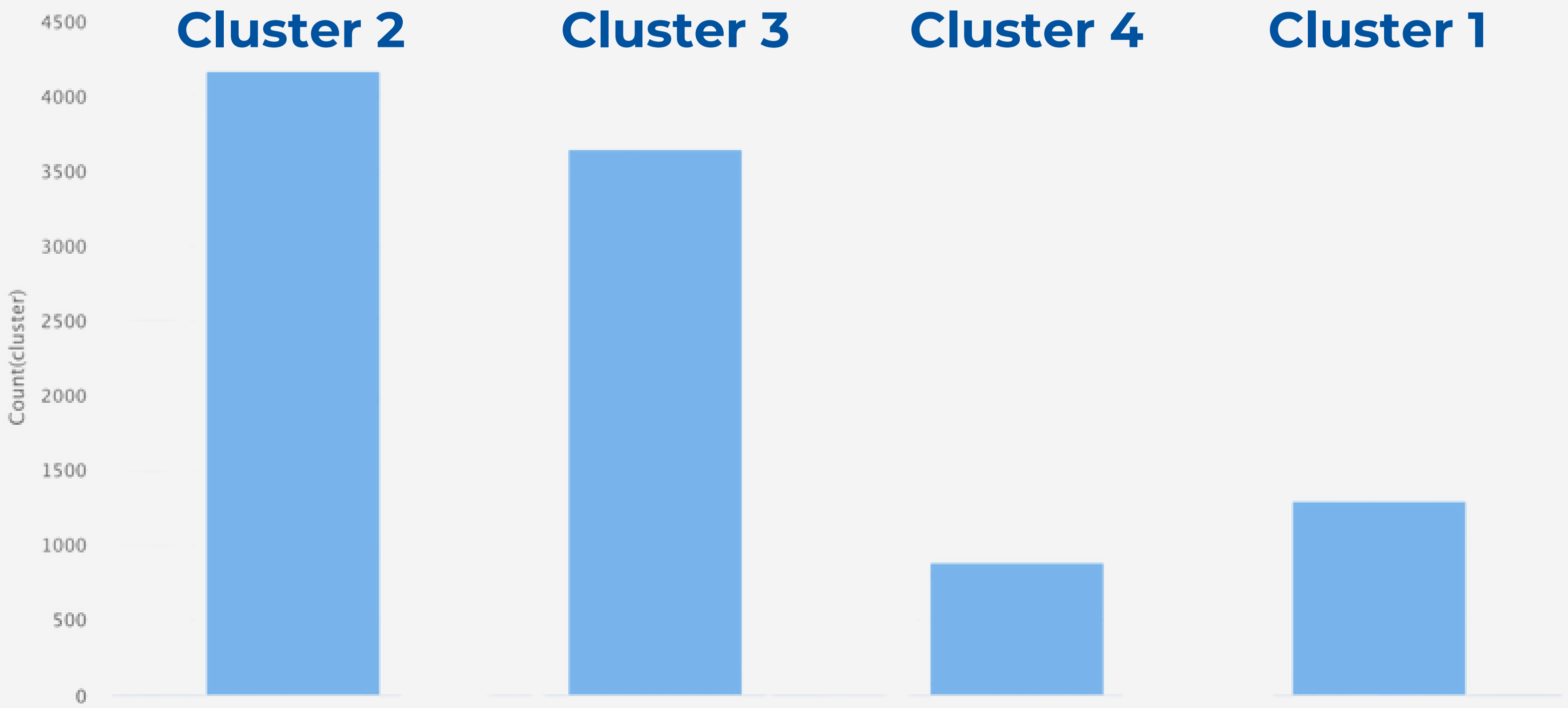
Filtering by Information Gain

Chose the top k attributes including diagnoses, age, race, and gender

Separate CSV File

Finalized csv file was used for patients that specifically were readmitted





Appendix G: Cluster Analysis

	Cluster 1	Cluster 2	Cluster 3	Cluster 4
	13.01% of Readmitted Patients	**41.64%** of Readmitted Patients	**36.48%** of Readmitted Patients	**8.87%** of Readmitted Patients
Gender	**100% Female**	**100% Female**	**100% male**	**100% male**
Race	86.55% African American	100% Caucasian	100% Caucasian	80.34% African American and 10% Hispanic
Primary Ages	22% in 70s 21.1% in 60s 21.6% in 50s 13% in 40s	Skews older since top age is: 70s (27.88%) Followed by 80s (23.37%)	29.41% is 70s 22.97% is 60s	24.8% in 60s 23% in 50s
Diagnoses	Respiratory Symptoms Heart Disease and Failure	Ischemic heart disease(4.7%) and heart failure (8.5%) Chronic bronchitis (3.6%)	Heart attacks, surgical complications, ischemic disease, heart failure	Pancreatitis and Heart Failure

Appendix G: Cluster Analysis

Pivot and cluster strategy: a preventive measure against diagnostic errors

Taro Shimizu [1,✉], Yasuharu Tokuda [2]

▸ Author information ▸ Article notes ▸ Copyright and License information

PMCID: PMC3508570 PMID: 23204855

Aunt Anne	Grandma Sally	Uncle Jerry	Mr. James
Diabetes in women **increases the risk of heart disease** more in women **than it does in men** because more likely to have added risk factors like obesity(3)	Diabetes relates to bronchitis because of chronic inflammation due to damaging blood vessels	Diabetes relates to bronchitis because of chronic inflammation due to damaging blood vessels	Pancreatis is related to pancreas being under so much strain from diabetes
High glucose levels from diabetes **damage blood vessels** that control your heart	Respiratory illnesses also may be risk factors for diabetes	High glucose levels from diabetes damage blood vessels that control your heart and pulmonary vessels	Men are more likely to be predisposed to heart attacks
Adults with diabetes are **twice as likely** to have **heart disease** or strokes (2).	Older women are more likely to develop heart failure because of a decline in estrogen (3) leads to more bad cholesterol	Adults with diabetes are twice as likely to have heart disease or strokes (2)	Adults with diabetes are twice as likely to have heart disease or strokes (2)

Sources

1. https://pmc.ncbi.nlm.nih.gov/articles/PMC4783289/#:~:text=Strategies%20that%20provide%20increased%20support,associated%20with%20lower%20readmission%20risk.
2. https://pmc.ncbi.nlm.nih.gov/articles/PMC6812156/#:~:text=1.,care%20for%20COPD%20(13).
3. https://pmc.ncbi.nlm.nih.gov/articles/PMC7211949/#:~:text=patients%20%5B85%5D.-,Treatment%20adherence,be%20vital%20in%20reducing%20readmissions.
4. https://www.saintlukeskc.org/health-library/discharge-instructions-chronic-bronchitis
5. https://www.sciencedirect.com/science/article/pii/S2589845022000719#:~:text=Decreasing%20readmission%20in%20acute%20pancreatitis&text=Sorrento%20et%20al.%20reported%20a,cholecystectomy%20and%20endoscopic%20biliary%20intervention).&text=Munigala%20et%20al.,Chu%20et%20al.&text=Ding%20et%20al.,-SNNAP%20Score&text=Whitlock%20et%20al.,-PASS%20Score&text=Buxbaum%20et%20al.,-Impact%20of%20readmission
6. https://pmc.ncbi.nlm.nih.gov/articles/PMC8418292/
7. https://hcup-us.ahrq.gov/reports/statbriefs/sb278-Conditions-Frequent-Readmissions-By-Payer-2018.jsp#:~:text=Bar%20chart%20that%20shows%20the,$16%2C400%20for%20privately%20insured%20stays.
8. https://www.ncbi.nlm.nih.gov/books/NBK606114/#:~:text=Understanding%20the%20root%20causes%20of,among%20patients%20with%20serious%20illnesses.
9. https://pmc.ncbi.nlm.nih.gov/articles/PMC8418292/
10. https://pmc.ncbi.nlm.nih.gov/articles/PMC4783289/#:~:text=Strategies%20that%20provide%20increased%20support,associated%20with%20lower%20readmission%20risk.
11. https://hbr.org/2014/10/the-value-of-keeping-the-right-customers
12. https://www.mdclarity.com/rcm-metrics/average-revenue-per-inpatient-admission#:~:text=their%20overall%20revenue.-,Average%20Revenue%20per%20Inpatient%20Admission%20Benchmark,and%20improve%20their%20financial%20performance.
13. https://www.mathematica.org/publications/national-trends-of-hospital-revenue-profit-and-labor-costs-2011-2022#:~:text=The%20national%20median%20Operating%20Profit,in%20operating%20expenses%20in%202022.
14. https://cost.sidecarhealth.com/c/endocrinologist-visit-cost
15. https://cost.sidecarhealth.com/c/nurse-visit-cost
16. https://cost.sidecarhealth.com/c/nurse-visit-cost
17. https://cost.sidecarhealth.com/c/pulmonologist-visit-cost
18. https://www.bls.gov/oes/current/oes291141.htm
19. https://www.smartpractice.com/shop/category?id=406575&m=SPM#:~:text=what%20to%20expect.-,100%20for%20$68.00%20%2D%20$0.68%20each,Reorder%20Products
20. https://www.ziprecruiter.com/Salaries/Hospital-Front-Desk-Clerk-Salary
21. https://www.mayoclinic.org/diseases-conditions/gallstones/diagnosis-treatment/drc-20354220#
22. https://www.goodrx.com/health-topic/diagnostics/ultrasound-cost-without-insurance
23. https://pmc.ncbi.nlm.nih.gov/articles/PMC4087585/#:~:text=About%20one%2Dthird%20of%20acute,history%20of%20chronic%20alcohol%20consumption.