

# Master of Applied Data Science Portfolio Presentation

*N'Dea Jackson  
Syracuse University  
Spring 2021*

---

# Introduction

The Applied Data Science program at Syracuse University is offered jointly by the School of Information Studies and the Martin J. Whitman School of Management. This Master of Applied Data Science degree program is designed to be a professional program of study, with a strong emphasis on the applications of data science to enterprise operations and processes, particularly in the areas of data capture, management, analysis, and communication for decision making.

The courses that were chosen to represent the learning objectives of the Masters of Applied Data Science Program were:

- IST 659: Database Administration
- IST 719: Data Visualization
- IST 707: Data Analytics
- MAR 653: Marketing Analytics

# Master of Applied Data Science Program Learning Objectives

1. Describe a broad overview of the major practices in data science
2. Collect and organize data
3. Identify patterns in data visualization, statistical analysis, and data mining
4. Develop alternative strategies based on their data
5. Develop a plan of actions to implement the business decisions derived from the analyses
6. Demonstrate communication skills regarding data and its analysis for managers, IT professionals, programmers, statisticians, and other relevant professionals in their organization
7. Synthesize the ethical dimensions of data science practice (e.g., privacy)



IST 659

# Database Administration

## Concepts and Management

---

# Jackson Memorial Hospital Database

**IST 659: Database Administration**

Learning Objectives Met:

- Collect/Create and organize data
- Develop alternative strategies for efficiency
- Develop a plan of action to implement the business decisions derived from analysis
- Demonstrate communication skills

Project overview:

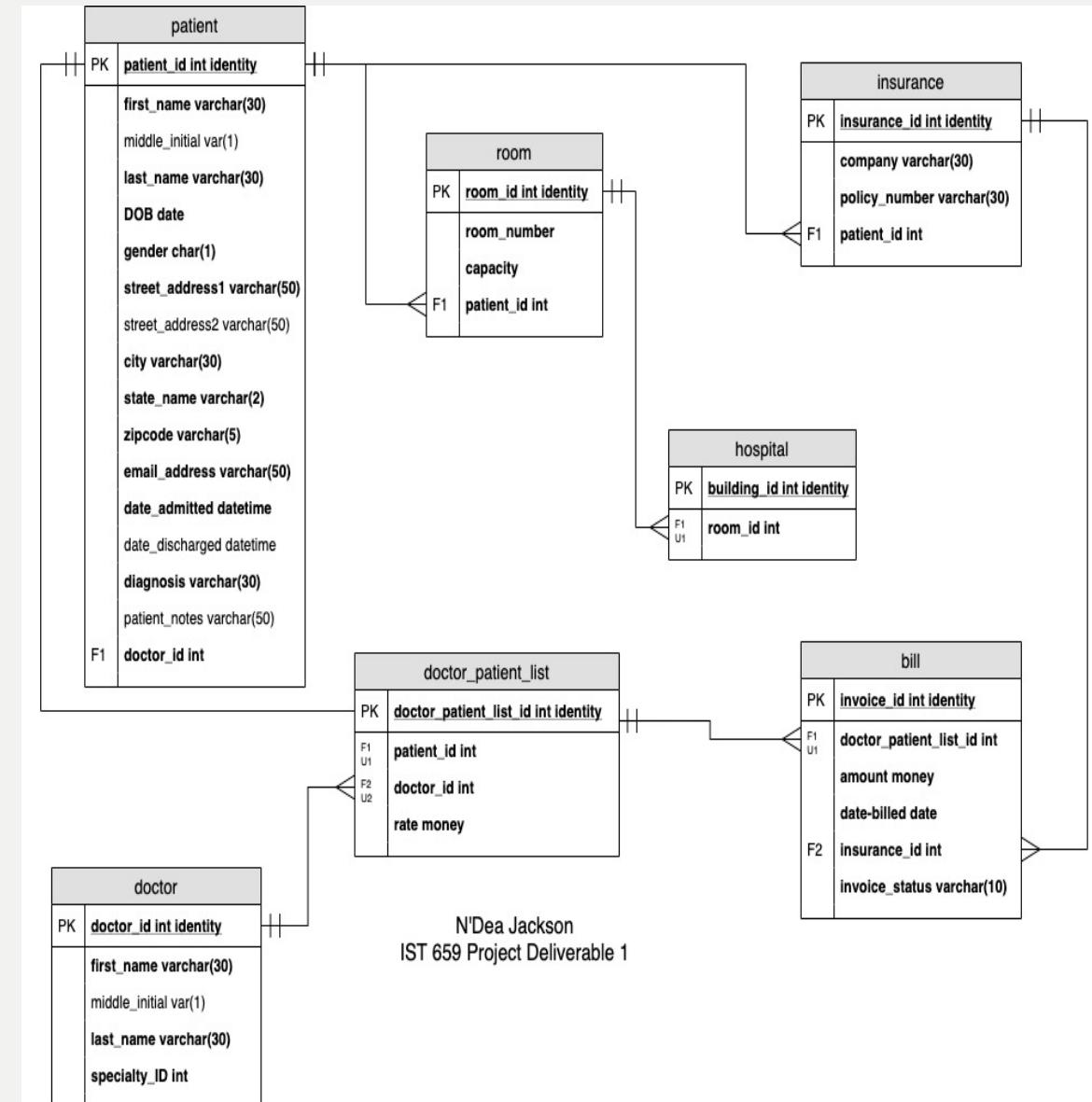
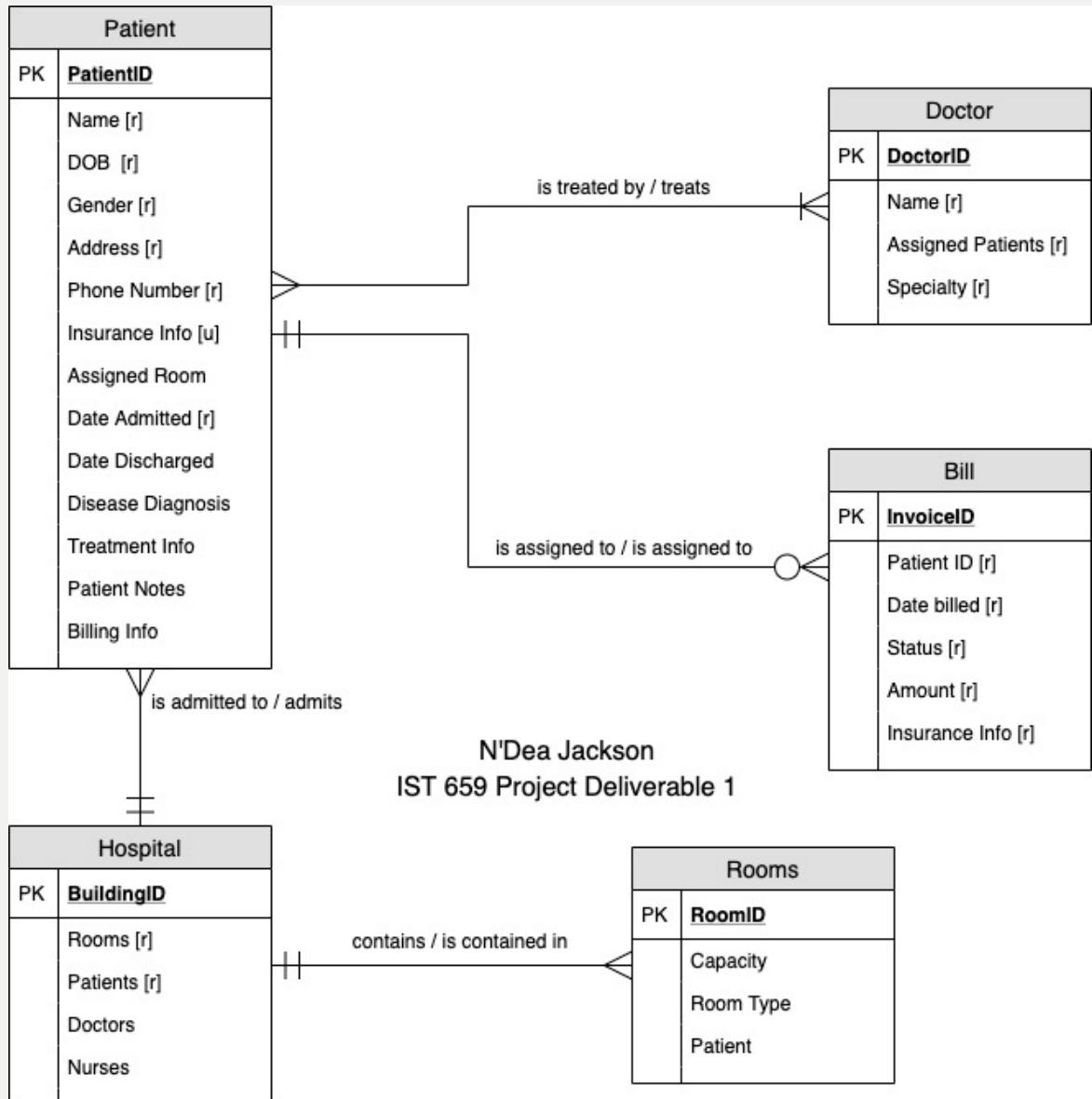
Under the direction of Professor Chad Harper, a hospital database was designed for a fictitious hospital in the hopes of:

- Providing the hospital with a better way to maintain all hospital records including:
  - Patient information
  - Doctor-Patient assignments
  - Patient room assignments
  - Patient invoice status

Overall benefit to the hospital:

- Minimize paperwork
- Eliminate siloed hospital systems
- Increase level of patient care and staff efficiency

Language used: SQL



Question 5: What room is Jermaine Cole patient assigned to?

```
SELECT * FROM PatientRoomAssignment WHERE PatientLastName = 'Cole'
```

	PatientFirstName	PatientMiddleInitial	PatientLastName	RoomNumber	RoomCapacity
1	Jermaine	L	Cole	412	1

Question 4: What is the status of Jason Webster's invoice?

```
SELECT * FROM PatientBillStatus WHERE InvoiceID = 8
```

	PatientFirstName	PatientMiddleInitial	PatientLastName	InvoiceID	InsuranceID	InvoiceAmount	InvoiceBillDate	InvoicePaidDate	invoice_status
1	Jason	NULL	Webster	8	109	2000.00	2020-03-04	NULL	UNPAID

Question 6: How old is Cody Carrington?

```
SELECT patient.patientID,  
       patient.first_name AS PatientFirstName,  
       patient.middle_initial AS PatientMiddleInitial,  
       patient.last_name AS PatientLastName,  
       (DateDiff(year, DOB, GETDATE())) AS AGE  
  FROM patient WHERE patient.last_name = 'Carrington'
```

	patientID	PatientFirstName	PatientMiddleInitial	PatientLastName	AGE
1	15	Cody	W	Carrington	70

Question 1: What patients are assigned to Mark Manson?

```
SELECT * FROM DoctorsAndPatients WHERE DoctorLastName = 'Manson'
```

	PatientFirstName	PatientMiddleInitial	PatientLastName	AssignedDoctor	DoctorFirstName	DoctorMiddleInitial	DoctorLastName
1	Kevin	NULL	Smith	6	Mark	NULL	Manson
2	Hilary	V	Banks	6	Mark	NULL	Manson
3	Cody	W	Carrington	6	Mark	NULL	Manson
4	Michaela	A	Nelson	6	Mark	NULL	Manson
5	William	D	Campbell	6	Mark	NULL	Manson

# Reflection

## IST 659: Database Administration

### Learning Objectives Met:

- Collect/Create and organize data
- Develop alternative strategies for efficiency
- Develop a plan of action to implement the business decisions derived from analysis
- Demonstrate communication skills

- Creating a database taught students the importance of relationship mapping as well as the importance of how data is stored.
- Missing or misrepresented relationships can dramatically change the effectiveness of the database
- Skills learned in this course are transferrable to the enterprise workforce as well as to future courses such as Advanced Database Management and Data Warehousing
- Project allowed students practice with identifying stakeholders, creating questions relevant to the stakeholders chosen, and creating business rules to govern the relationships that existed.



# IST 719

# Data Visualization

---

# Boston Crime Data Report

IST 719: Data Visualization

Learning Objectives Met:

- Collect and organize data
  - Going a step further to identify patterns in the data via visualization and statistical techniques
- Develop a plan of actions to implement the business decisions derived from the analyses
- Demonstrate communication skills regarding the data

Project Overview:

Under the direction of Professor Gary Krudys, a poster visualization was designed in order to display the most prevalent crimes that occurred in Boston neighborhoods.

Overall benefit of the visualization:

- Boston Police Department (BPD) would gain a better understanding of which neighborhoods could benefit from an increased police presence
- City officials would gain a better understanding of programs they could place in certain areas to curb criminal activity
- Making this data public would allow residents, potential residents, and visitors to be more aware of their surroundings

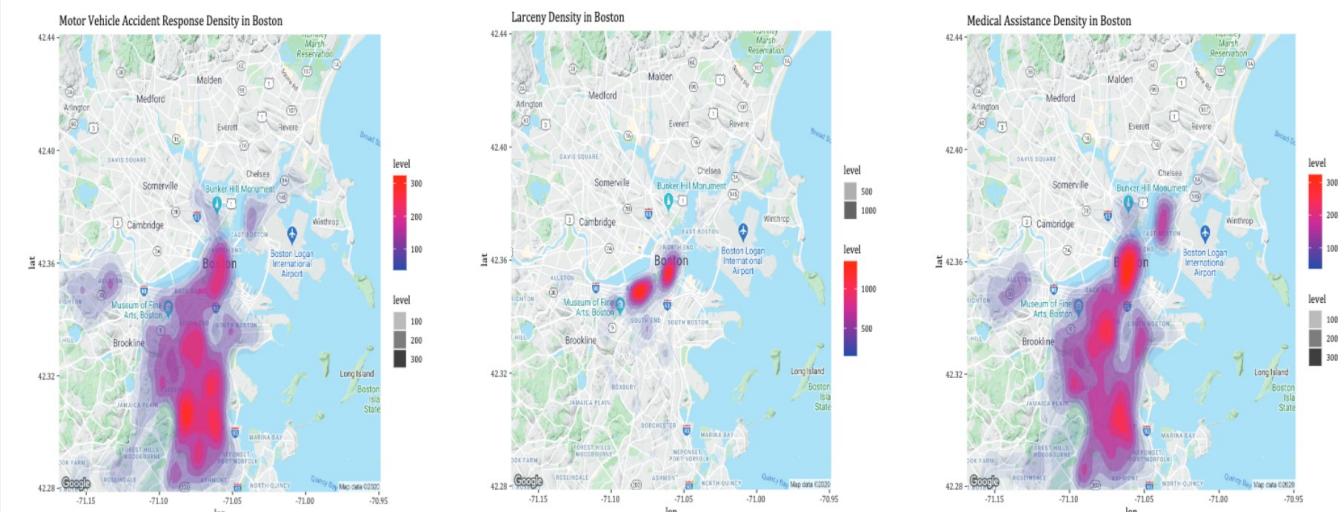
Language used: R

# The Data

## IST 719: Data Visualization

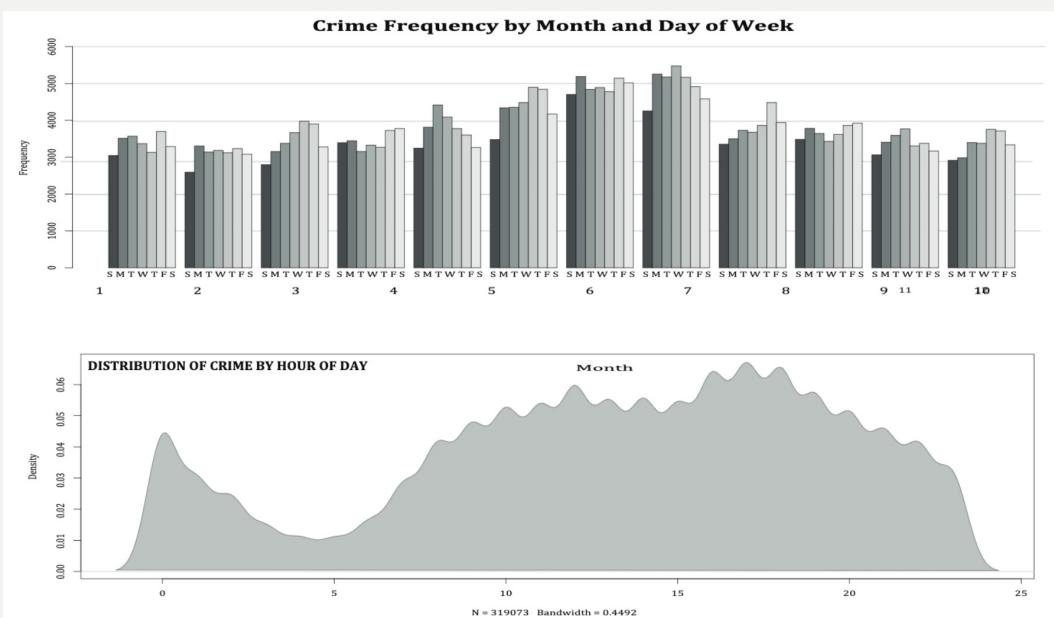
What crimes occur most frequently?

- Heat maps were created of the three most frequently occurring crimes that occur in Boston neighborhoods:
  - Motor Vehicle Accident Response
  - Larceny (including shoplifting, bicycles, purse snatches)
  - Medical Assistance (including sick/injured/medical/suicide attempt/sudden death/death investigation)



When do crimes occur?

- Histograms and density plots were created to examine the frequency of crimes over the course of the four years being analyzed:
  - 48,496 incidents were recorded on Fridays, making it the day of the week with the highest crime frequency
  - Over the course of a 24-hour period, crime appears to drop after midnight and steadily rise until it reaches its peak at 5PM.
  - Yearly, the summer months, May-August, crime rises.



# Reflection

## IST 719: Data Visualization

### Learning Objectives Met:

- Collect/Create and organize data
- Develop alternative strategies for efficiency
- Develop a plan of action to implement the business decisions derived from analysis
- Demonstrate communication skills

- This course allowed students the space to gain practical knowledge on how to use R to perform data cleaning and preparation on a wide range of datasets
- Taught how to gain insights from data as well as how to effectively communicate findings through visualizations
  - Create visualizations that don't compromise the data being presented
  - Accurately and effectively portray data
- Sharpened communication skills by creating data stories and tailoring presentations to specific audiences
- Skills learned in this course were transferrable to other courses throughout the program and are also extremely valuable in the workforce.



# IST 707

# Data Analytics

---

# National Basketball Association Shot Analysis

IST 707: Data Analysis

Learning Objectives Met:

- Collect and organize data
- Identify patterns in the data via visualization and data mining
- Develop alternative strategies based on the data used
- Implement business decisions derived from the analyses
- Demonstrate communication skills by reporting on findings with recommendations

Project Overview:

Under the direction of Professor Jeremy Bolton, various data mining, data preparation, and other relevant techniques were used to analyze shots that were taken by National Basketball Association (NBA) players.

Overall benefit of the analysis:

- Help the league improve performance by predicting things such as whether a game would be won based on several factors
- Drive fan satisfaction ratings in the tumultuous 2019-2020 season

*Language used: R*

# Methodology

## About the data

- NBA dataset that was scraped from the NBA's REST API before it was made publicly unavailable.
- The NBA uses a 6-camera system that can track 2-dimensional player locations 25 times per second.
- After all data preparation, there were 281 unique player IDs represented across 1808 unique games.

## Methods and Results

### Methods used:

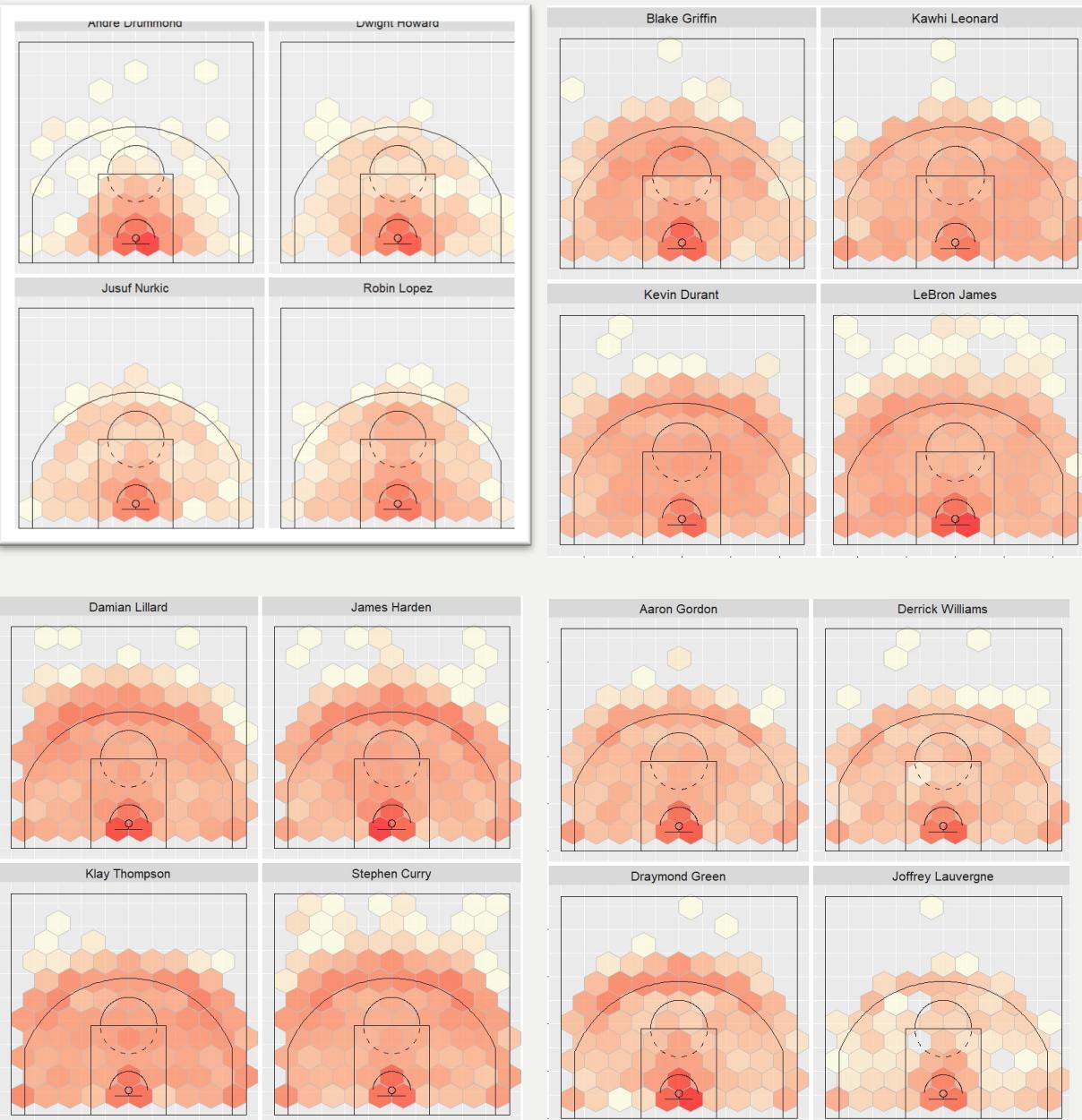
- Decision trees, association rule mining, support vector machines
  - SVM had the highest accuracy of predicting whether a shot would be made with 79.3%
  - Decision trees proved to have the highest accuracy of predicting whether a game would be won with 64.54%

# Methodology

Clustering was used in this project to determine the different player types that were represented within the data.

Four shooter types:

- Lay-up range
- Elbow range
- 3-Point range
- Combination player



# Reflection

## IST 707: Data Analysis

### Learning Objectives Met:

- Collect and organize data
- Identify patterns in the data via visualization and data mining
- Develop alternative strategies based on the data used
- Implement business decisions derived from the analyses
- Demonstrate communication skills by reporting on findings with recommendations

- This course allowed students the opportunity to test various techniques to solve the same problem.
- Highlights for students the strengths and weaknesses of various modeling techniques
- Getting different results from different techniques shows the importance of testing multiple models before making business decisions
- Using multiple solutions to solve the same problems can provide businesses with the most precise results and save costs down the line



# MAR 653

# Marketing Analytics

---

# San Francisco Airport Data: Pandemic Recovery Survey

MAR 653: Marketing Analytics

Learning Objectives Met:

- Collect and organize data
- Identify patterns in the data via visualization and data mining
- Develop alternative strategies based on the data used
- Implement business decisions derived from the analyses
- Demonstrate communication skills by reporting on findings with recommendations

Project Overview:

Under the direction of Professor Shaam Ramamurthy, marketing strategies and quantitative analysis techniques were applied to the San Francisco Airport Survey dataset to determine which customers SFO should target for traveling improvements.

Overall benefit of the analysis:

- Help the San Francisco Airport (SFO) determine which travelers may need more encouragement before feeling safe flying
- Make recommendations to SFO to improve the passenger travel experience during the pandemic

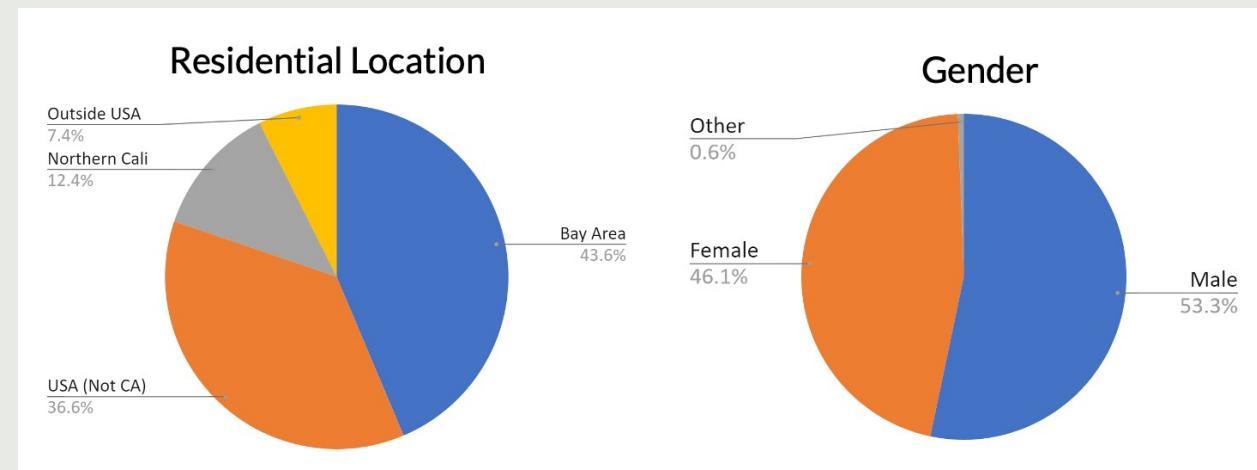
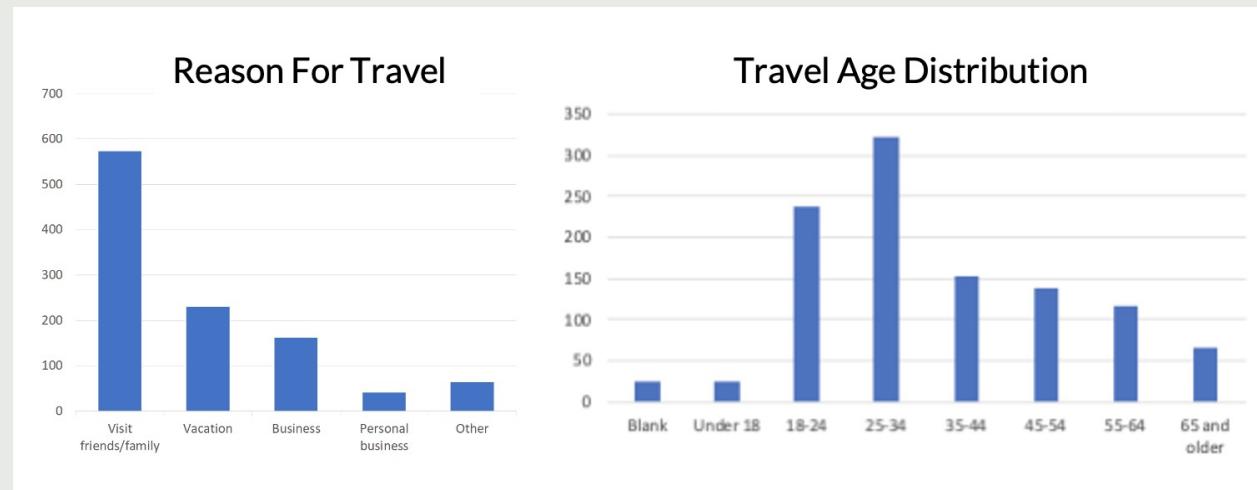
*Tools used: Python, Excel, Orange*

# Methodology

## MAR 653: Marketing Analytics

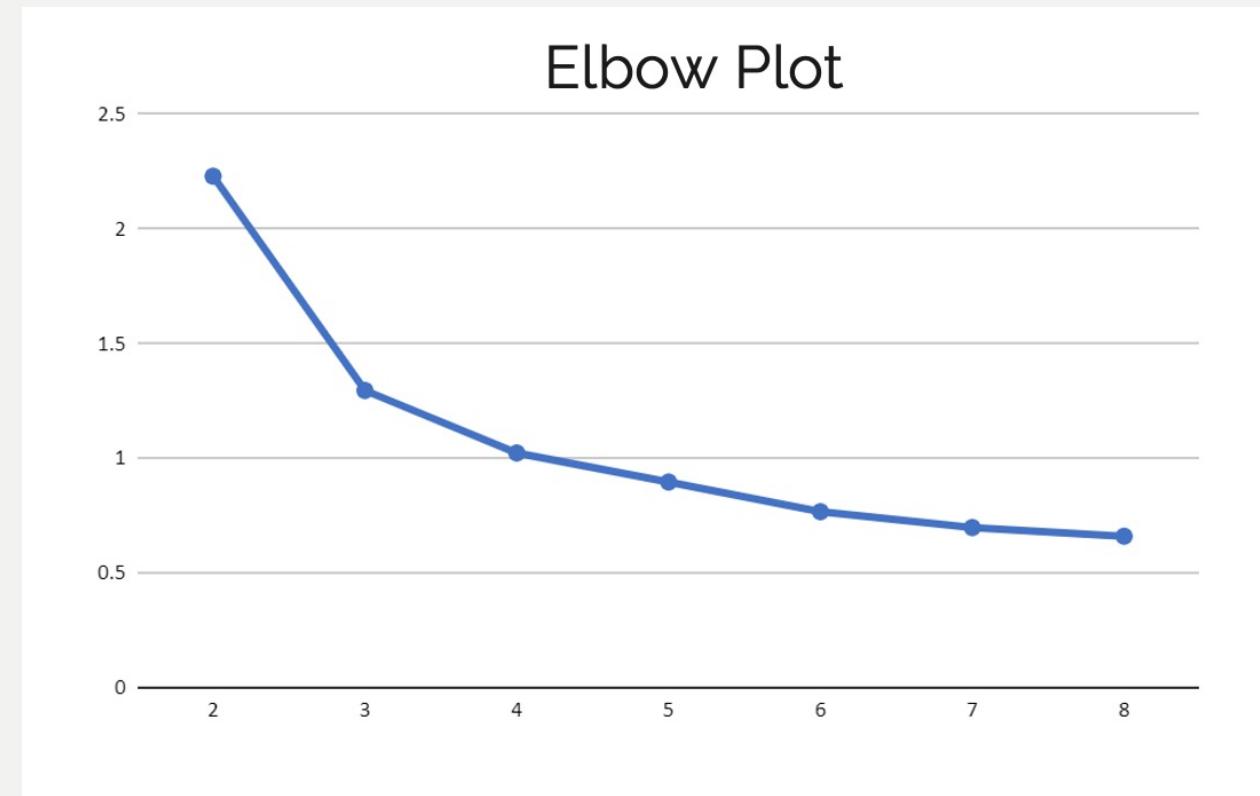
### About the data

- Surveys were conducted on paper surveys within the SFO airport
- 1,086 participants
- 16 questions that collected information on traveler demographics and traveler comfortability.
- Used by SFO to gauge satisfaction levels and travel preferences of travelers in hopes to remain above satisfactory



# Methodology

## MAR 653: Marketing Analytics



- K-Means clustering used to segment customers into groups that could be used by SFO. 3 clusters used.
- Most important factors for travelers:
  - **47%** of travelers selected **strict social distancing guidelines and enforcement.**
  - **14%** of travelers selected **having hand sanitizer readily available.**
  - **11%** of travelers selected **ensuring everyone was wearing a mask.**

## Concerned Travelers (1)

These people tended to express more concern about traveling during the pandemic, and safety precautions.

- Are “very concerned” about flying during the pandemic
- Most likely to arrive at the airport by private transportation
- Lowest average income of all three groups
- More business travel than other groups
- Oldest average age of all three groups

## Apprehensive Travelers (2)

These people expressed moderate concern about flying, but were the least satisfied about SFO’s current safety precautions

- Gave SFO lowest rating on overall response to protect travelers
- Lowest ratings on all questions about SFO’s current precautions
- Lowest average age
- Least likely to be traveling for business
- Highest average income

## Confident Travelers (3)

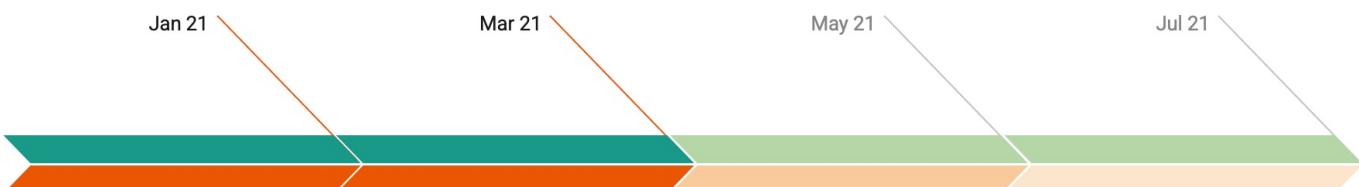
These people felt comfortable traveling during the pandemic, and expressed little concern about the airport’s current safety precautions.

- Most vacation travelers
- Highest average income
- Most likely to arrive by public transportation
- Highest rating on overall response to protect travelers precautions
- Most likely to indicate “Covid is a hoax/overblown”
- Most likely to want mask removal for vaccinated travelers

# Recommendation

MAR 653: Marketing Analytics

## 8 Month Plan: Recommendations



### Phase 1: Immediate

- Enforce mask policies and social distance guidelines.
- Installing additional barriers for workers.
- Readily accessible sanitation stations.

### Phase 2: Assurance

- Transparency into process of sanitising planes and public areas.
- Improved management of airport food services- enforce grab and go.

### Phase 3: Check-in

### Phase 4: Adapt

- Issue another survey to understand current needs.
- Adjust policies according to vaccinations (opening maskless vax sections).

- Respond to survey questions by adjusting policy

### Which customers to target:

- Very concerned about flying
- Low ratings on current airport policies
- Older people
- Less likely to use public transportation

### Why:

- Alleviate tensions associated with resuming post pandemic travel.
- Have SFO take extra precautions to meet the needs of the most vulnerable groups of people.

# Reflection

## MAR 653: Marketing Analytics

### Learning Objectives Met:

- Collect and organize data
- Identify patterns in the data via visualization and data mining
- Develop alternative strategies based on the data used
- Implement business decisions derived from the analyses
- Demonstrate communication skills by reporting on findings with recommendations
- This course allowed students the opportunity collect and organize data of their choosing to target a specific customer set and optimize some marketing goal.
- Students were taught how to create various strategies to lead to the same outcome, improving client lifetime value.
- Customer segmentation helps students to learn their audiences in hopes to provide more value.
- Communication skills were further developed through group presentations

# Recap of Learning Objectives

This portfolio presentation shows the successful implementation and understanding of the learning objectives of the Masters of Applied Data Science program. The courses taken in this program compounded on each other and gradually increased in complexity.

1. Describe a broad overview of the major practices in data science
2. Collect and organize data
3. Identify patterns in data visualization, statistical analysis, and data mining
4. Develop alternative strategies based on their data
5. Develop a plan of actions to implement the business decisions derived from the analyses
6. Demonstrate communication skills regarding data and its analysis for managers, IT professionals, programmers, statisticians, and other relevant professionals in their organization
7. Synthesize the ethical dimensions of data science practice (e.g., privacy)

# Conclusion

- The Applied Data Science program at Syracuse University is designed to be a professional program of study with a strong emphasis on the applications of data science to enterprise operations and processes, particularly in the areas of data capture, management, analysis, and communication for decision making.
- The skills that students learn in the program provided them with many multifaceted approaches to dealing with both structured and unstructured data.
- Students gained practical experience using several different languages and tools to solve real business problems including *R*, *Python*, *Excel*, *SQL*, and *Microsoft Access*.
- Through several means including reports, oral presentations, poster creation, and this portfolio, students learned communication skills valuable for presenting to stakeholders and other specific audiences.
- Through courses, students gained an understanding of both security and ethical boundaries in data.



# Thank you!