



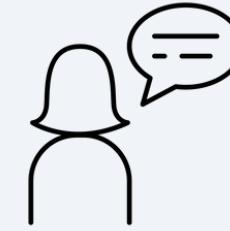
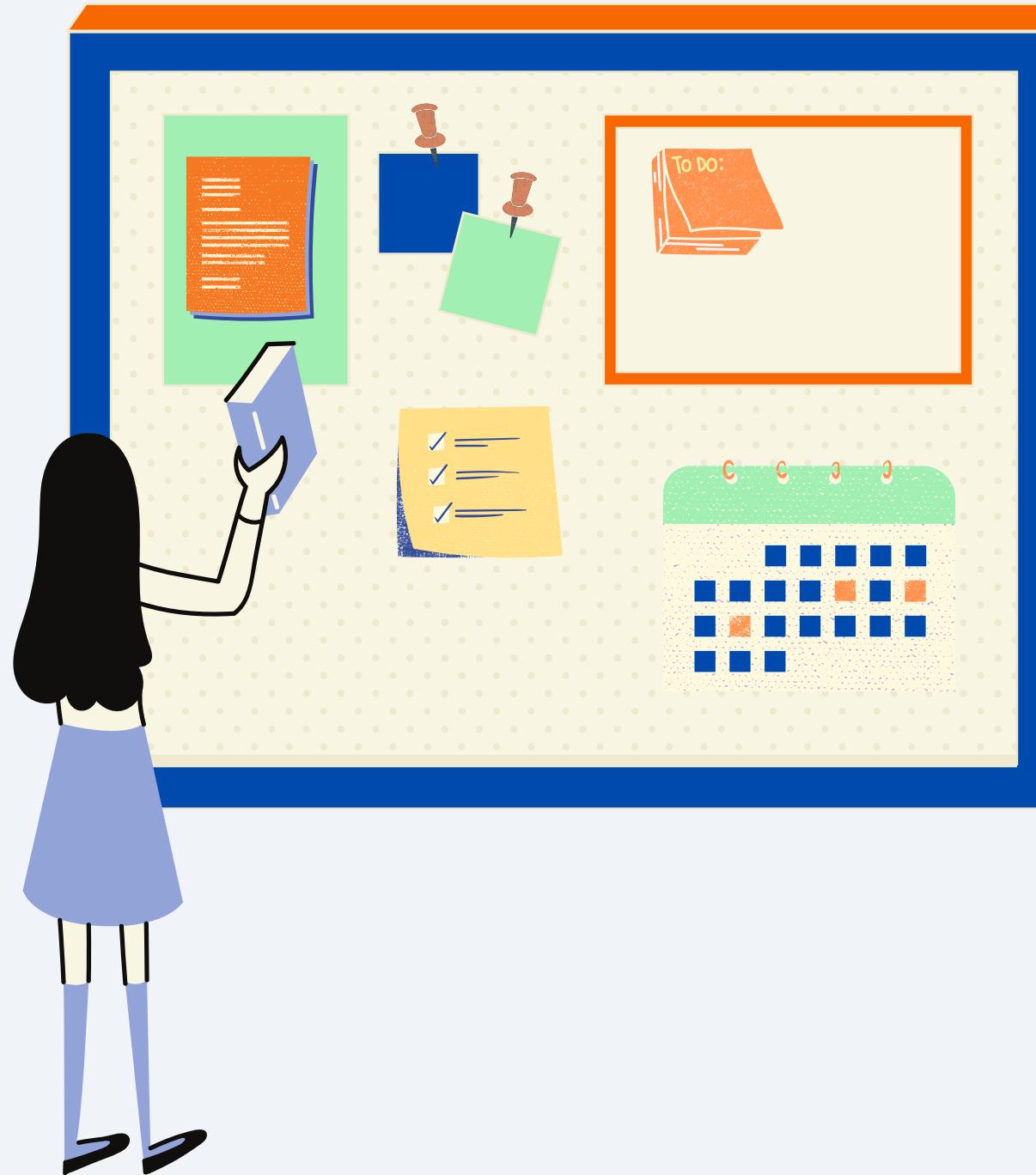
SYRACUSE UNIVERSITY
SCHOOL OF INFORMATION STUDIES

MS APPLIED DATA SCIENCE PORTFOLIO MILESTONE

RAYANNA HARDUARSINGH
FALL 2021



AGENDA



INTRODUCTION

Self-Introduction & Background
Overview of ADS Program Learning Goals



COURSES & PROJECT DELIVERABLES

IST 687 - Introduction to Data Science
IST 659 - Data Admin Concepts & Database Management
IST 719 - Data Visualization



LEARNING GOALS

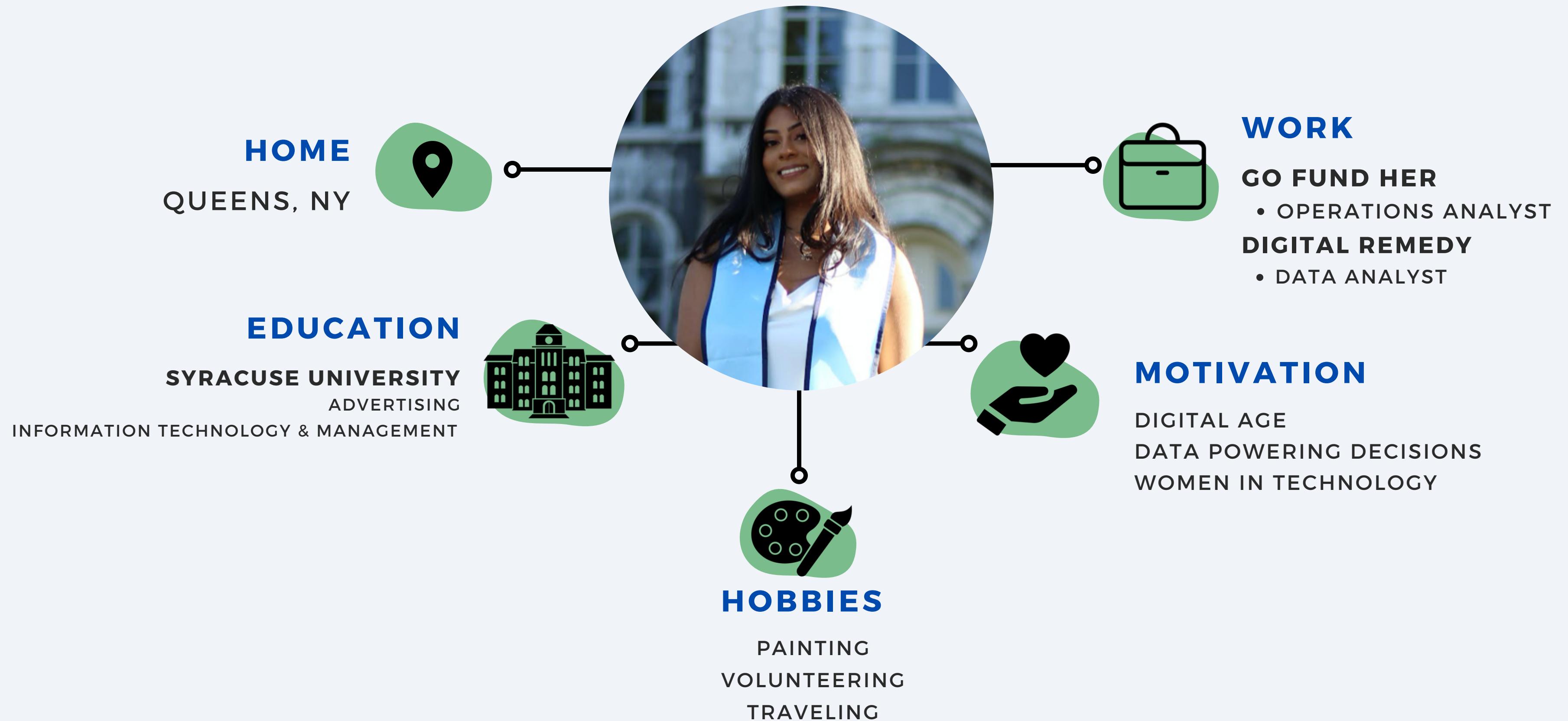
Demonstration & achievement of each learning goal



CONCLUSION

Self-Reflection & Future Plans

INTRODUCTION



LEARNING GOALS



1. COLLECT, STORE, AND, ACCESS DATA BY IDENTIFYING AND LEVERAGING APPLICABLE TECHNOLOGIES
2. CREATE ACTIONABLE INSIGHT ACROSS A RANGE OF CONTEXTS USING DATA AND THE FULL DATA SCIENCE LIFE CYCLE
3. APPLY VISUALIZATION AND PREDICTIVE MODELS TO HELP GENERATE ACTIONABLE INSIGHT
4. USE PROGRAMMING LANGUAGES SUCH AS R AND PYTHON TO SUPPORT THE GENERATION OF ACTIONABLE INSIGHT
5. COMMUNICATE INSIGHTS GAINED VIA VISUALIZATION AND ANALYTICS TO A BROAD RANGE OF AUDIENCES
6. APPLY ETHICS IN THE DEVELOPMENT, USE AND EVALUATION OF DATA AND PREDICTIVE MODELS

COURSES & PROJECT DELIVERABLES



IST 687: INTRODUCTION TO DATA SCIENCE

COURSE DESCRIPTION

- Led by Professor Jeffrey Saltz
- Data collecting, processing, transformation, management, and analysis.
- Utilized “R” to conduct applied statistics, data visualization, text mining, and machine learning.



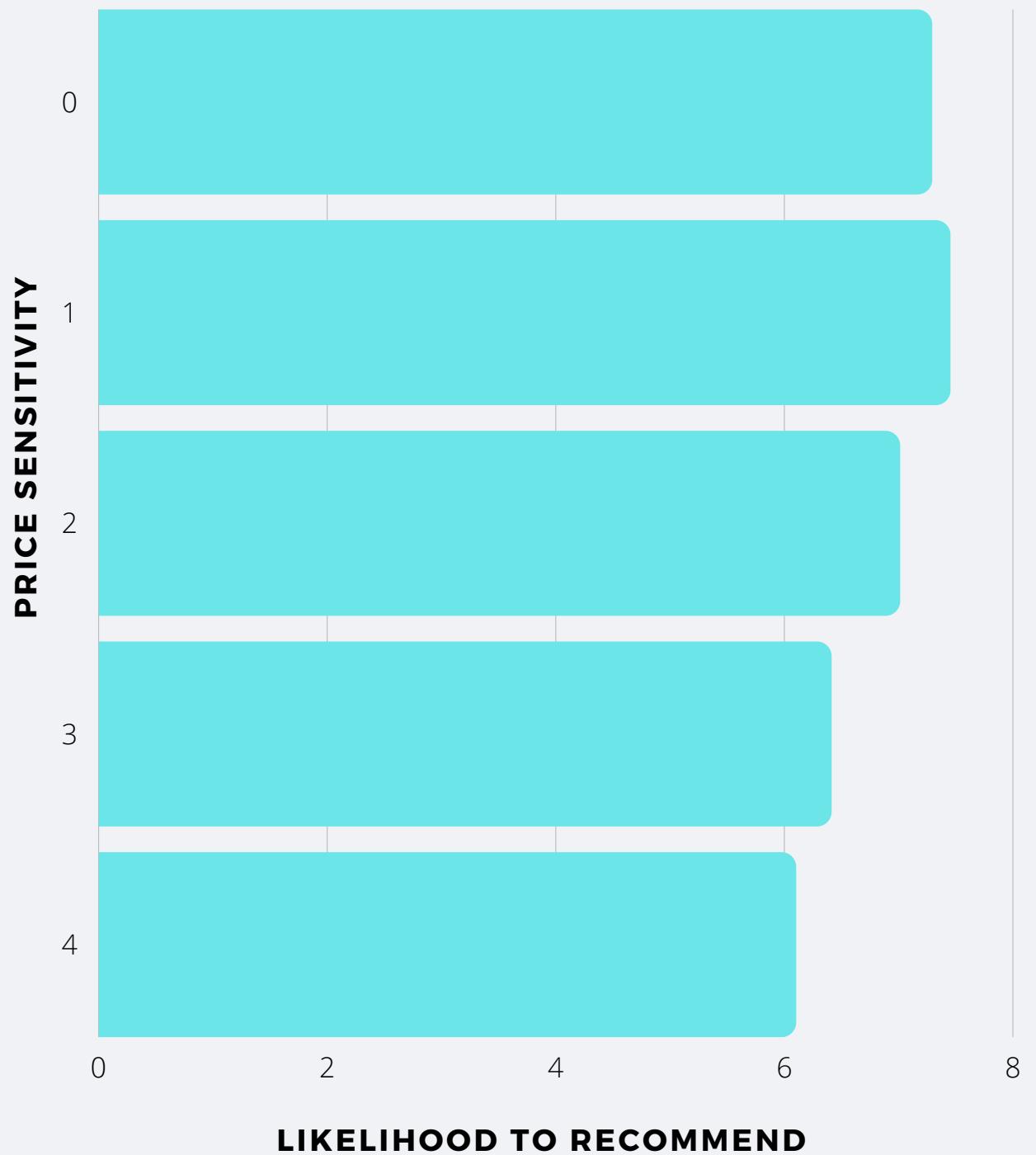
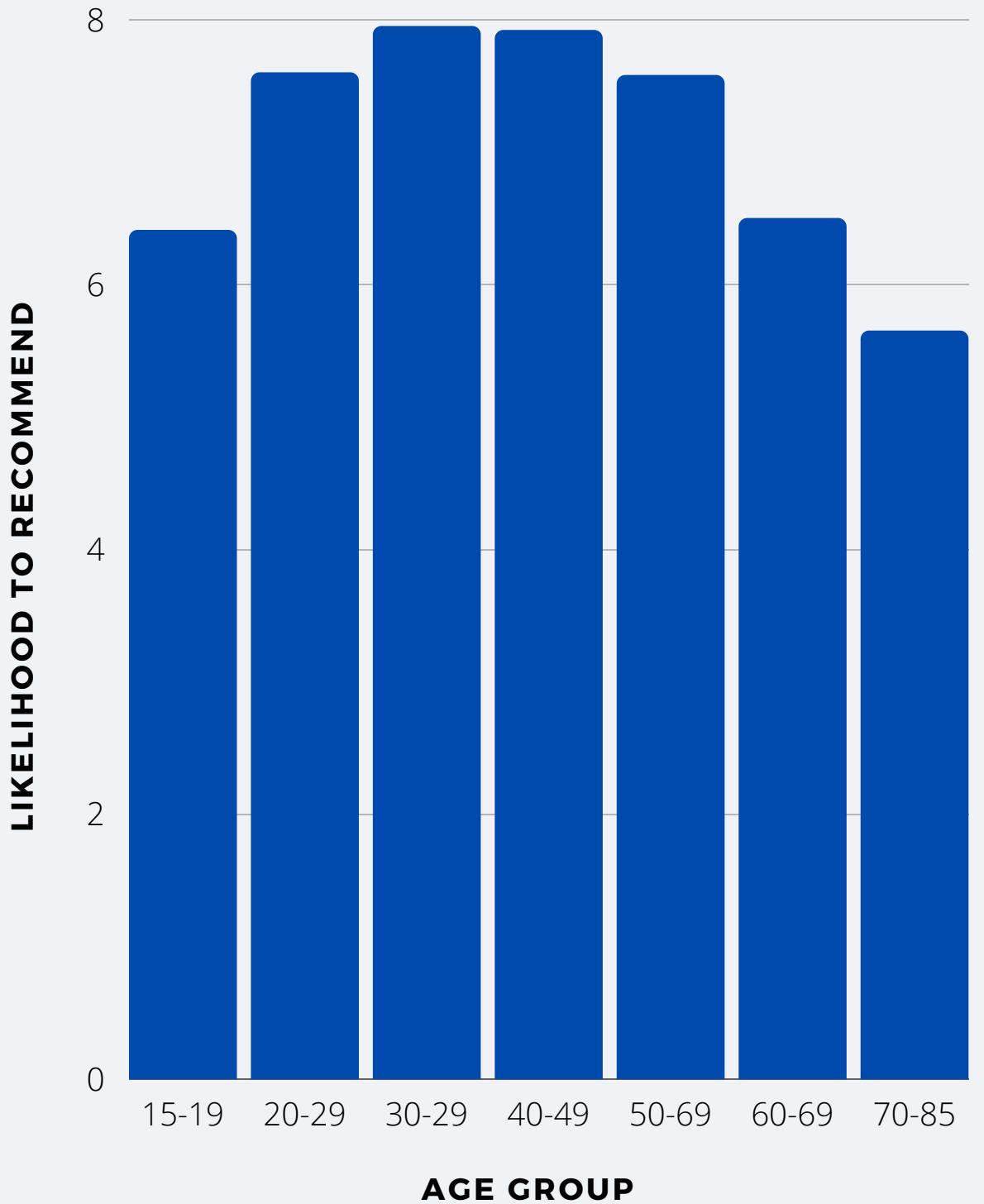
ABSTRACT

- A survey with 88,000 responses was conducted by Southeast Airlines travelers who rated their overall satisfaction on a scale of 1-10
- Improve Southeast Airline's customer churn by exploring, analyzing, and interpreting results
- Evaluated significant factors to predict which key factors have an influence on the satisfaction score.
- Discovered important trends in the data and provided Southeast Airlines with business recommendations to increase customer satisfaction.

DATA EXPLORATION

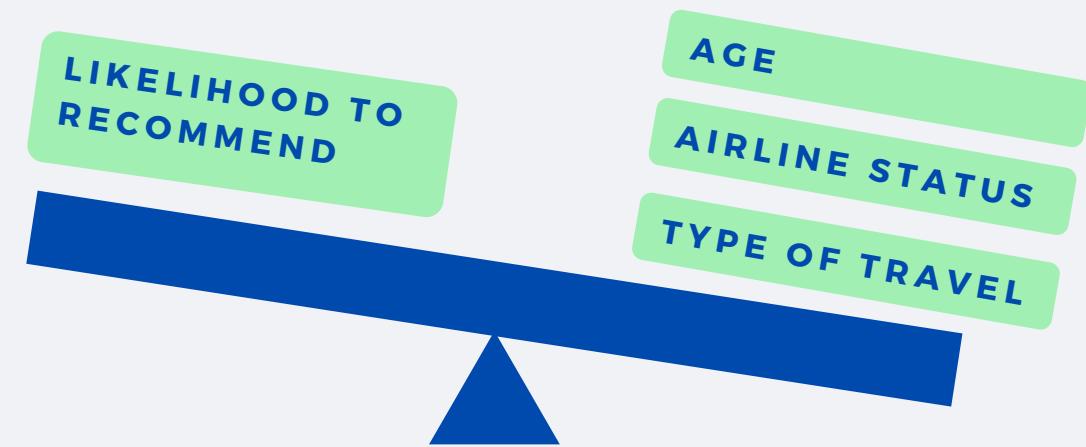
DESCRIPTIVE STATS

- Average rating of 8.99
- 60,000 travelers were of Blue Airline Status
- Over 60,000 flights were for business purposes



1.

LINEAR REGRESSION



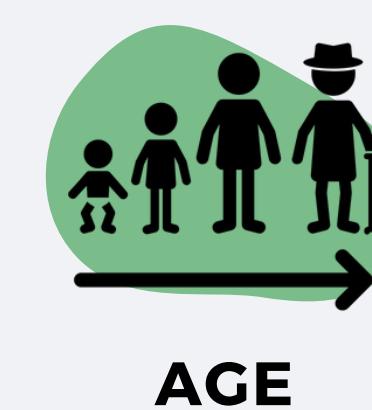
3.

ASSOCIATION RULE MINING



2.

SVM



AGE

FLIGHT DELAY

PRICE

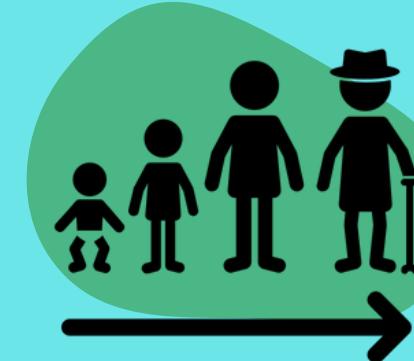
4.

TEXT MINING



80 WORDS OF "DELAY"
OUT OF 100 WAS
MENTIONED

ACTIONABLE INSIGHTS & RECOMMENDATIONS



AGE 60-85

CUSTOMERS AGED BETWEEN 60-85 TENDED TO GIVE LOWER RATINGS

- Senior citizens should be extra-cared for such as offering blankets/pillows on their flights and how can better accommodate them to enjoy their trip.



**PRICE
SENSITIVITY =1**

RATING APPEARED TO BE GREATER WHEN THE PRICE IS LOW.

- Offer seasonal promotions during special times of the year (Thanksgiving, Christmas, Summer, etc.).
- Point based system to attract price sensitive demographics encourage them to keep flying with us



FLIGHT DELAY

LOW RATINGS WHEN WAITING MORE THAN 5 MINS FOR FLIGHT

- Access to VIP lounge
- Food/Snack vouchers or gift cards that passenger can avail in flight or at the airport(sent digitally)
- Complimentary snack

LEARNING GOALS APPLIED

1. CREATE ACTIONABLE INSIGHT ACROSS A RANGE OF CONTEXTS USING DATA AND THE FULL DATA SCIENCE LIFE CYCLE

- Collected, cleaned, and prepared a data set and recognized a problem (how can we reduce customer churn?)
- Performed descriptive statistics and visualizations to undertake exploratory data analysis, which lead to our prediction models.
- Tested our prediction models and created actionable insights to generate business recommendations.

2. APPLY VISUALIZATION AND PREDICTIVE MODELS TO HELP GENERATE ACTIONABLE INSIGHT

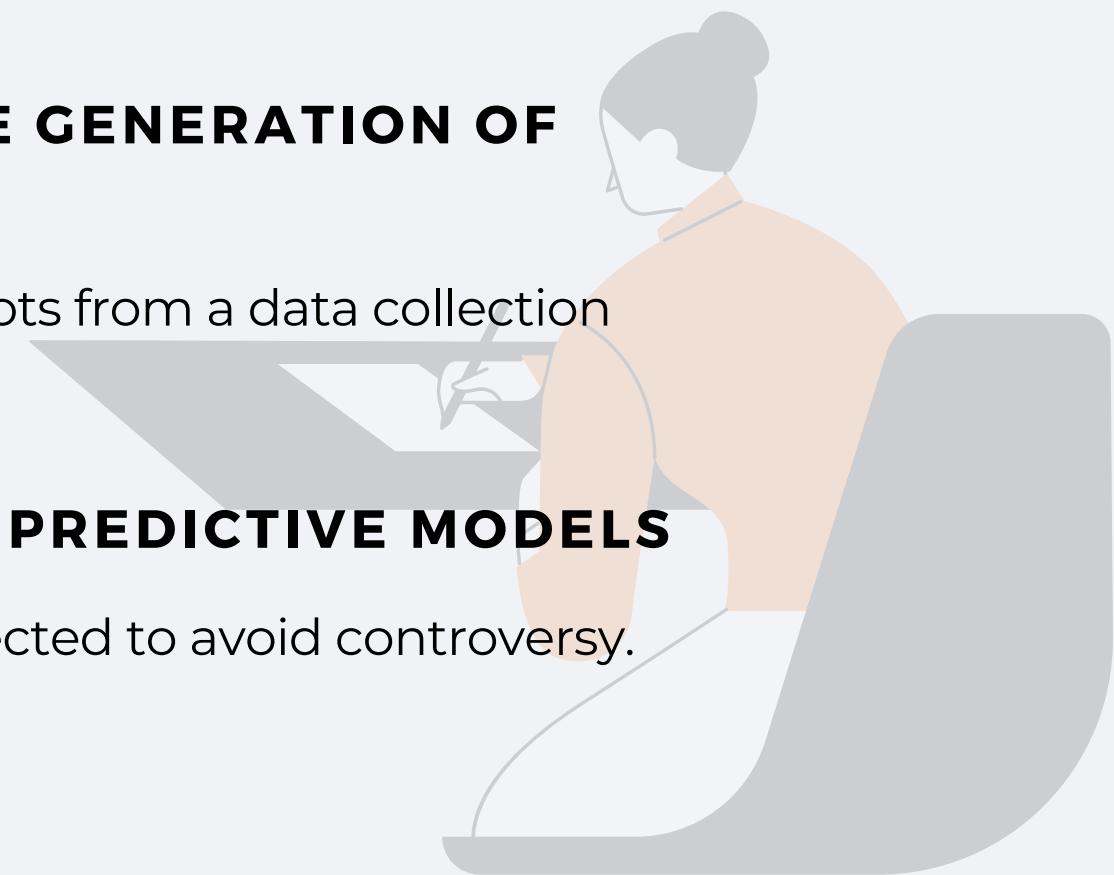
- Predictive analysis is performed with machine learning models that helps generate possible predictions for the future results
- Predictive models such as linear regressions and machine learning models may assist in the identification of relationships between two or more variables
- Identified correlations and found which major factors influence each other the most.
 - We witnessed, for example, how a traveler's bad rating is influenced by delayed flights.

3. USE PROGRAMMING LANGUAGES SUCH AS R AND PYTHON TO SUPPORT THE GENERATION OF ACTIONABLE INSIGHT

- Predictive modeling, machine learning methodologies, text mining, and creating visual plots from a data collection
- extracting insights and informations from a data set using its results

4. APPLY ETHICS IN THE DEVELOPMENT, USE AND EVALUATION OF DATA AND PREDICTIVE MODELS

- It's critical to consider how public perceptions of your findings and suggestions will be reflected to avoid controversy.
- To avoid public bias, we purposely avoided focusing on the gender variable.



IST 659: DATA ADMINISTRATION CONCEPTS & DATABASE MANAGEMENT

COURSE DESCRIPTION

- Led by Professor Michael Fudge
- Examine data structures, organize data, and implement data analysis
- Utilized Structured Query Language (SQL) to construct, model, and maintain databases as well as create queries.

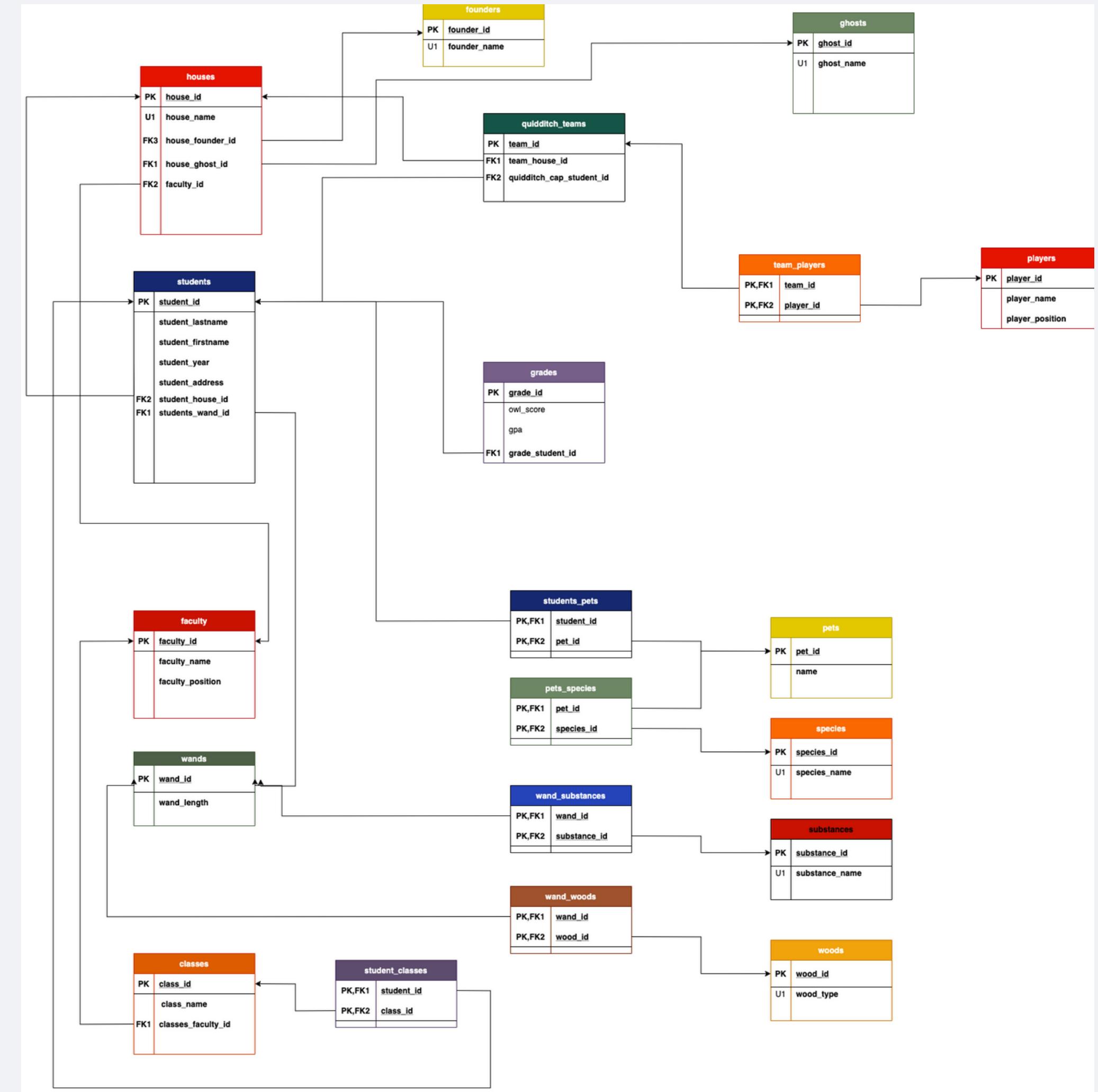


ABSTRACT

IMPLEMENT A FUNCTIONAL SYSTEM WITH A DATABASE

- Hogwarts School of Witchcraft and Wizardry is entering the digital era and need assistance in digitizing and managing data about their school, students, and staff from a technological aspect in order to improve organizational efficiency.
- Design a database management system for Hogwarts that would allow professors to efficiently manage and organize their institution's many departments in terms of students enrolled, classes, sports, and more.

LOGICAL MODEL



USER STORIES



Classes

As a faculty member, I should be able to see what classes a student is in so that they are on the right track to graduate.



Teacher

As a student, I should be able to see what teacher is teaching a class so that I can make my schedule appropriate to my liking.



Wands

As a student, I should be able to see the length and components of my wand so that if it ever goes missing I can know how to replace it.



Addresses

As a faculty member, I should be able to view student addresses in order to mail school correspondence so that students are prepared before the term starts.



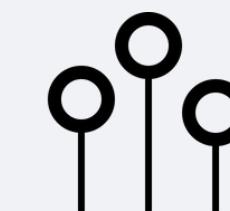
Pets

As a faculty member, I should be able to view the names and species of the pets each student has and will bring to the school so the number of pets are not exceeded.



Quidditch

As a faculty member, I should be able to view the quidditch players so that I can assign the correct number of players to a team.



```
--As a faculty member, I should be able to see what classes a student is in so that they are on the right track to graduate.
select students.student_id,
       student_firstname + ' ' + student_lastname as student_name,
       student_year,
       classes.class_id,
       classes.class_name
  from students
 join student_classes on students.student_id=student_classes.student_id
 join classes on student_classes.class_id=classes.class_id
```

[Results](#) [Messages](#)

	student_id	student_name	student_year	class_id	class_name
1	1	Harry Potter	5	1	Defense Against the Dark ...
2	1	Harry Potter	5	3	Astronomy
3	1	Harry Potter	5	4	Herbology
4	1	Harry Potter	5	8	Divination
5	1	Harry Potter	5	9	Care of Magical Creatures
6	2	Hermione Granger	5	1	Defense Against the Dark ...
7	2	Hermione Granger	5	3	Astronomy
8	2	Hermione Granger	5	4	Herbology
9	2	Hermione Granger	5	6	Transfiguration
10	2	Hermione Granger	5	8	Divination
11	2	Hermione Granger	5	9	Care of Magical Creatures
12	3	Ronald Weasley	5	1	Defense Against the Dark ...



EXTERNAL DATA MODEL

USER STORY #1

CLASSES

As a faculty member, I should be able to see what classes a student is in so that they are on the right track to graduate.





EXTERNAL DATA MODEL

```
--As a student, I should be able to see what teacher is teaching a class so that I can make my schedule appropriate to my liking.  
select class_id, class_name, faculty.faculty_id,  
faculty_firstname + ' ' + faculty_lastname as Professor_name  
from classes  
join faculty on classes.faculty_id = faculty.faculty_id
```

	class_id	class_name	faculty_id	Professor_name
1	1	Defense Against the Dark ...	2	Severus Snape
2	2	Potions	11	Horace Slughorn
3	3	Astronomy	9	Gilderoy Lockhart
4	4	Herbology	8	Pomona Sprout
5	5	Dark Arts	17	Alastor Moody
6	6	Transfiguration	3	Minerva McGonagall
7	7	Music	10	Filius Flitwick
8	8	Divination	15	Sybill Trelawney
9	9	Care of Magical Creatures	4	Rubeus Hagrid
10	10	Charms and Spells	5	Dolores Umbridge
11	11	History of Magic	6	Remus Lupin
12	12	Flying	12	Rolanda Hooch
13	13	Muggle Studies	14	Cuthbert Binns

USER STORY #2

TEACHER

As a student, I should be able to see what teacher is teaching a class so that I can make my schedule appropriate to my liking.





EXTERNAL DATA MODEL

```
--As a student, I should be able to see the components of my wand so that if it ever goes missing I can know how to replace it.  
select student_firstname + ' ' + student_lastname as student_name,  
wands.wand_id,wand_length,wood_type,substance_name from wands  
join wand_woods on wands.wand_id = wand_woods.wand_id  
join wand_substances on wands.wand_id = wand_substances.wand_id  
join woods on wand_woods.wood_id = woods.wood_id  
join substances on wand_substances.substance_id = substances.substance_id  
join students on wands.wand_id = students.wand_id
```

	student_name	wand_id	wand_length	wood_type	substance_name
1	Harry Potter	1	11.00	Elm	Dragon Heartstring
2	Hermione Granger	2	10.75	Ash	Phoenix Feathers
3	Ronald Weasley	3	12.00	Hazel	Unicorn Hair
4	Ginerva Weasley	4	9.50	Red Oak	Veela Hair
5	Fred Weasley	5	10.25	English Oak	Thestral Tail Hair
6	George Weasley	6	10.25	Sugar Maple	Troll Whisker
7	Draco Malfoy	7	10.00	Pine	Kelpie Hair
8	Luna Lovegood	8	11.25	Snakewood	Thunderbird Tail Feather
9	Neville Longbottom	9	12.25	Chestnut	Wampus Cat Hair
10	Padma Patil	10	9.00	Reed	White River Monster Spine

USER STORY #3

WAND

As a student, I should be able to see the length and components of my wand so that if it ever goes missing I can know how to replace it.

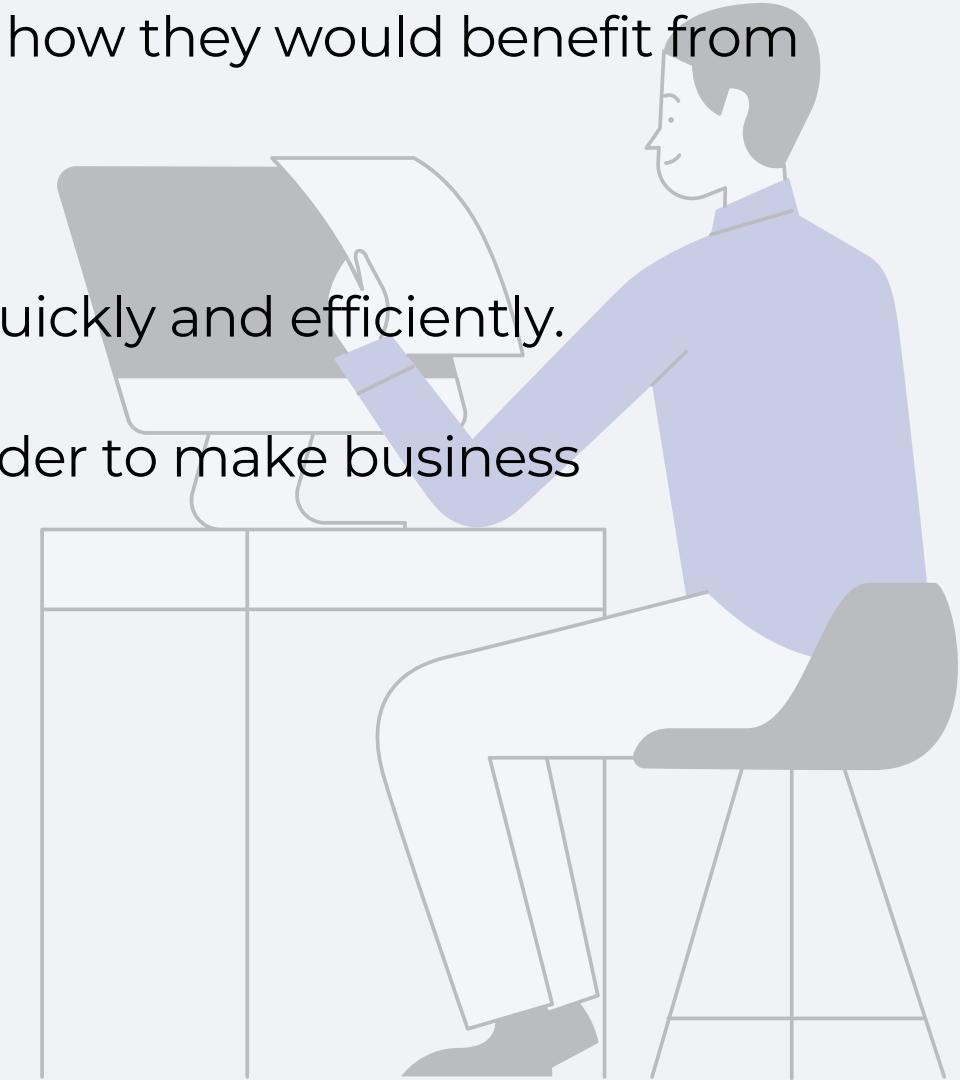


LEARNING GOALS APPLIED

COLLECT, STORE, AND, ACCESS DATA BY IDENTIFYING AND LEVERAGING APPLICABLE TECHNOLOGIES

1.

- Creating data from scratch to collect and store it into a database management system using SQL.
- Create multiple user stories or scenarios in which a user would access the system and how they would benefit from this data being transformed into a database.
- Learned how to properly clean and organize data so that analysis can be completed quickly and efficiently.
- Data management has taught me the importance of storing and accessing data in order to make business decisions and automate operational processes.



IST 719: DATA VISUALIZATION

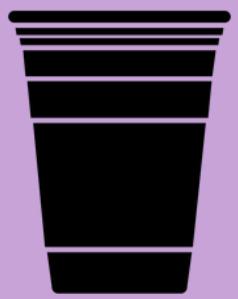
COURSE DESCRIPTION

- Led by Professor Jeff Hemsley
- Practiced a variety of skills and techniques for creating informational data visualizations
- Utilized RStudio's graphics environment to clean data, create custom plots, and visually explore data.
- Utilized Adobe Illustrator to apply numerous design concepts that visually communicated a story told from the data.



ABSTRACT

- Create multiple visualizations that tell a story of how alcohol consumptions among students affect their academic activities.
- Underage drinking has been a serious long-term issue in the United States and all over the world.
 - About 7.05 million Americans between the ages of 12 and 20 reported current alcohol consumption (2019 National Survey on Drug Use and Health).
 - Obtained student data from the University of California Irvine (UCI) that includes various social, gender, and study data
 - Included students' grade reports, age, alcohol consumption, mother's education, and more.



DOES UNDERAGE DRINKING AFFECT A STUDENT'S ACADEMIC PERFORMANCE?



VERY LIGHT DRINKER



LIGHT DRINKER



MEDIUM DRINKER

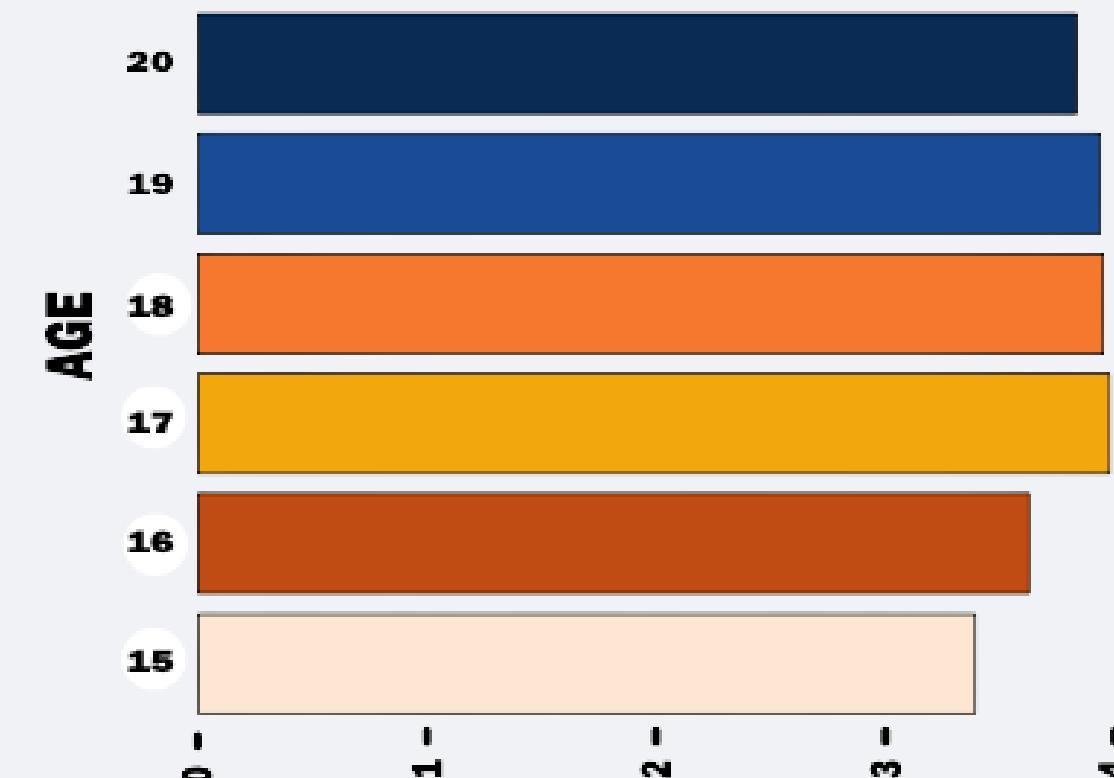


HEAVY DRINKER

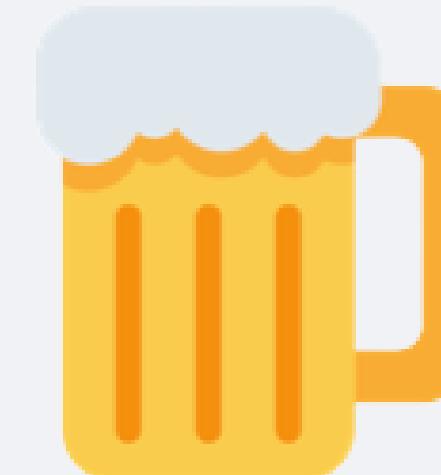
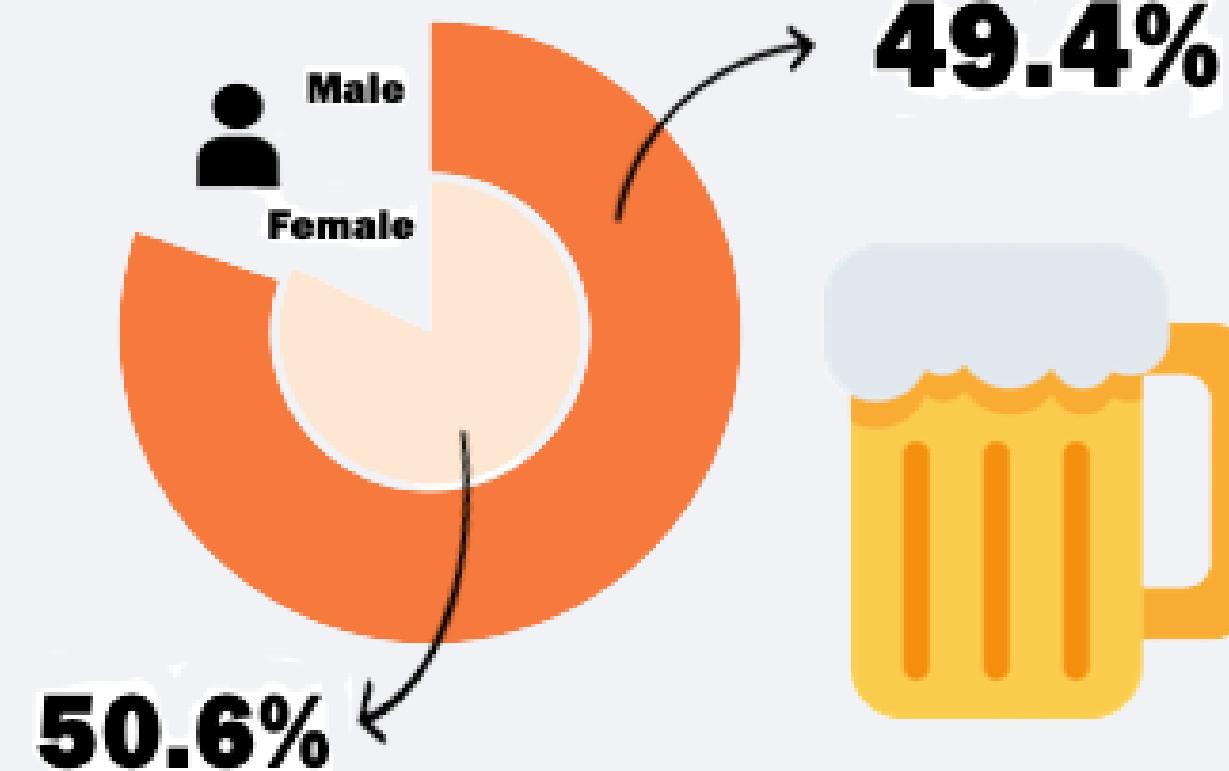


VERY HEAVY DRINKER

HOW OFTEN DO THEY DRINK?



NUMBER OF TIMES PER WEEK



STUDENT ALCOHOL CONSUMPTION

Underage drinking has been a serious long-term issue in the United States. It remains a huge concern as these behaviors can lead to negative consequences and effects. With multiple variables such as the amount of times a student studies, goes out, drink as well as their grades, students can see the effect alcohol has on them.

Author: Rayanna Harduarsingh

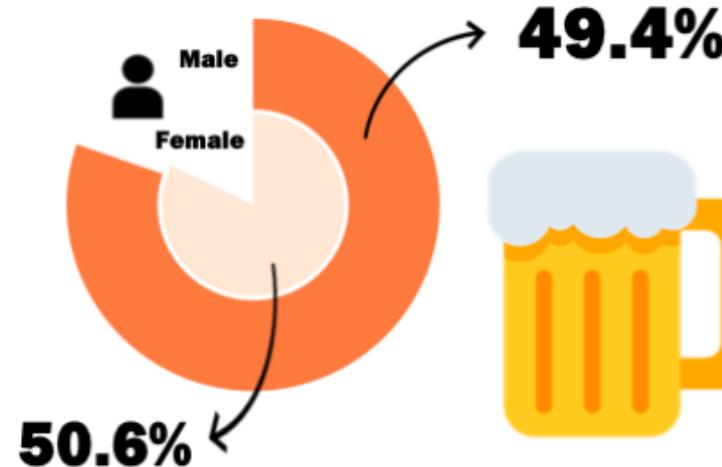
Class: IST 719 M001

Date: May 11, 2021

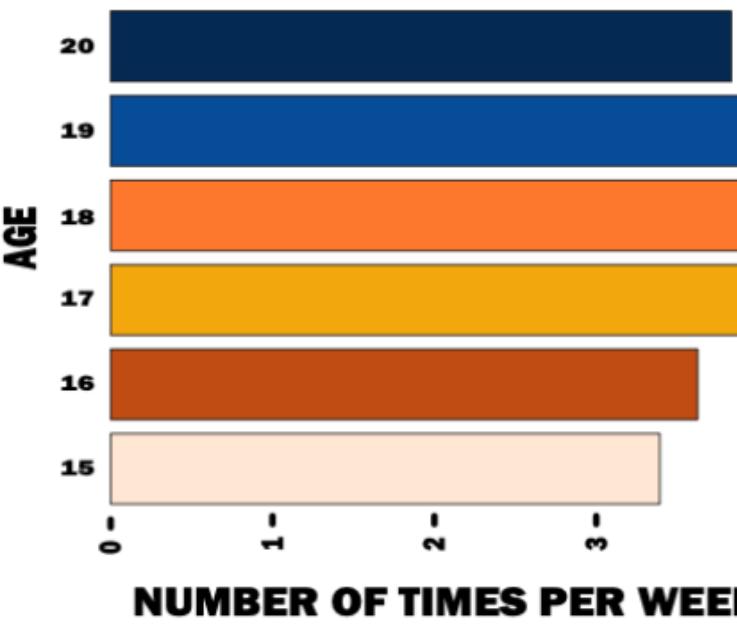
DATA DESCRIPTION

The University of California obtained student data from a survey on a Portuguese language course in secondary school. With 39 columns and 649 rows, the data set included various social, gender, and study data on these students as well as their drinking behaviors. The data combines weekday and weekend alcohol consumption among students which will be used to compare to their academic activities and general demographics.

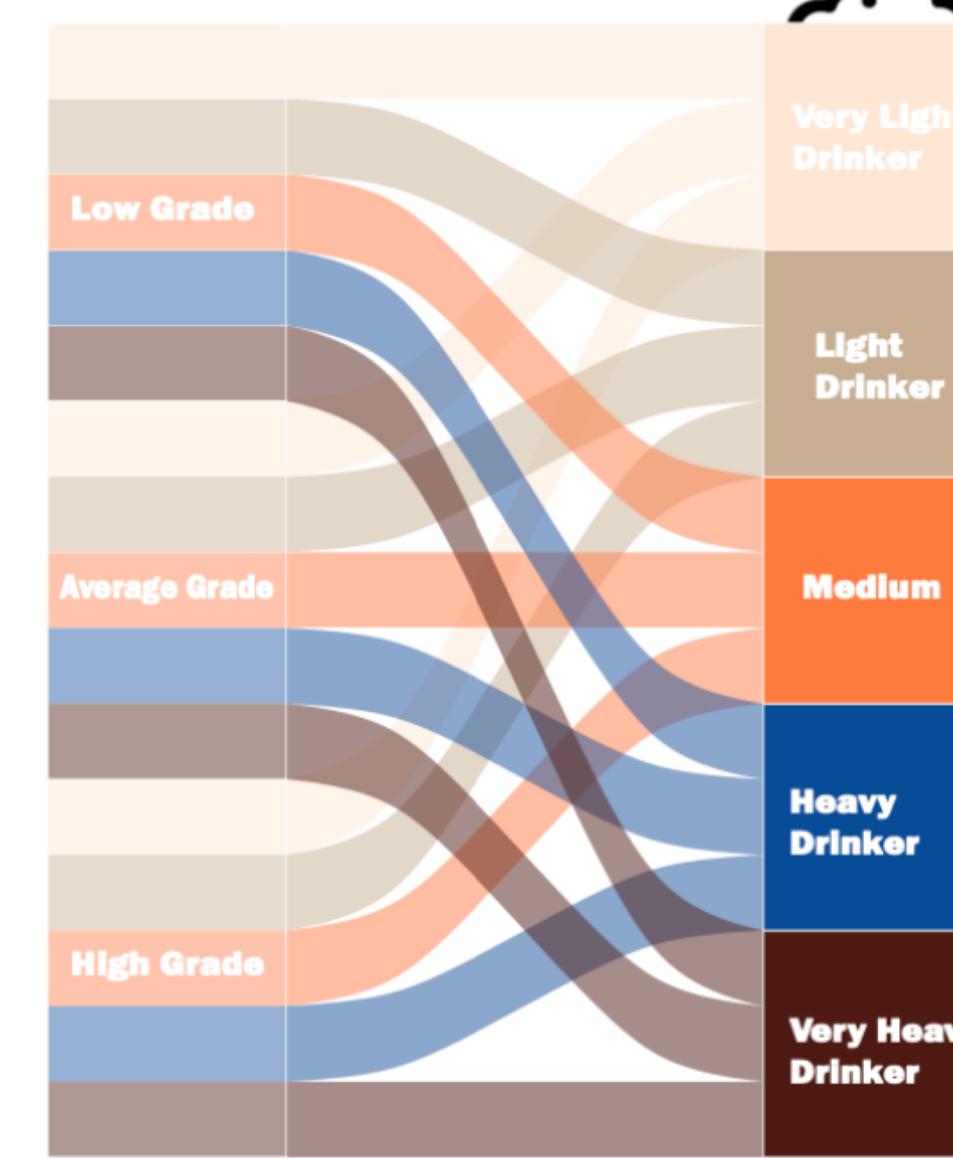
STUDENTS & UNDERAGE DRINKING



HOW OFTEN DO THEY DRINK?

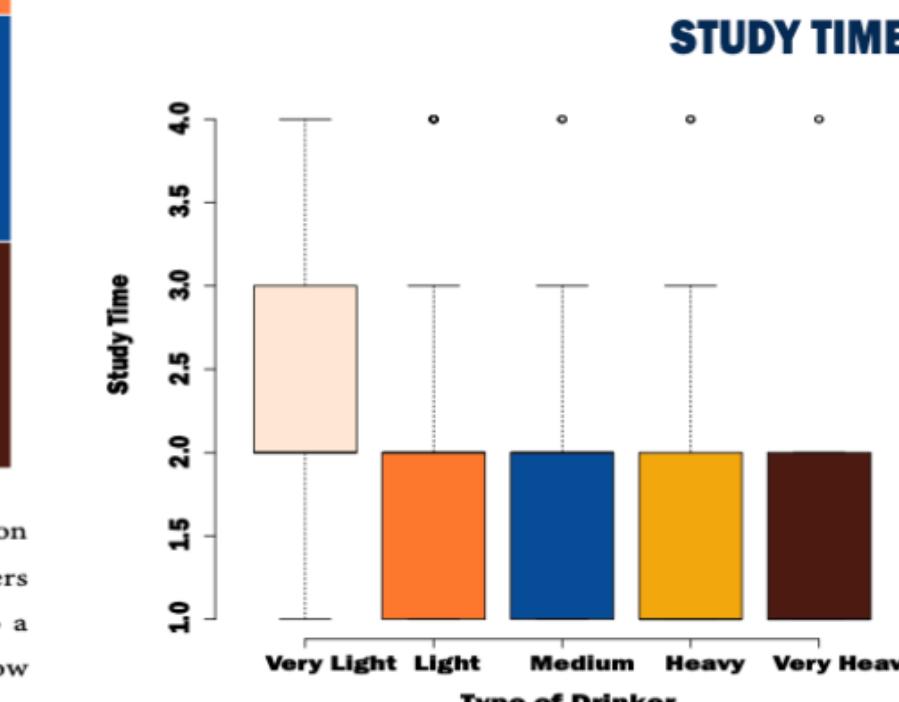
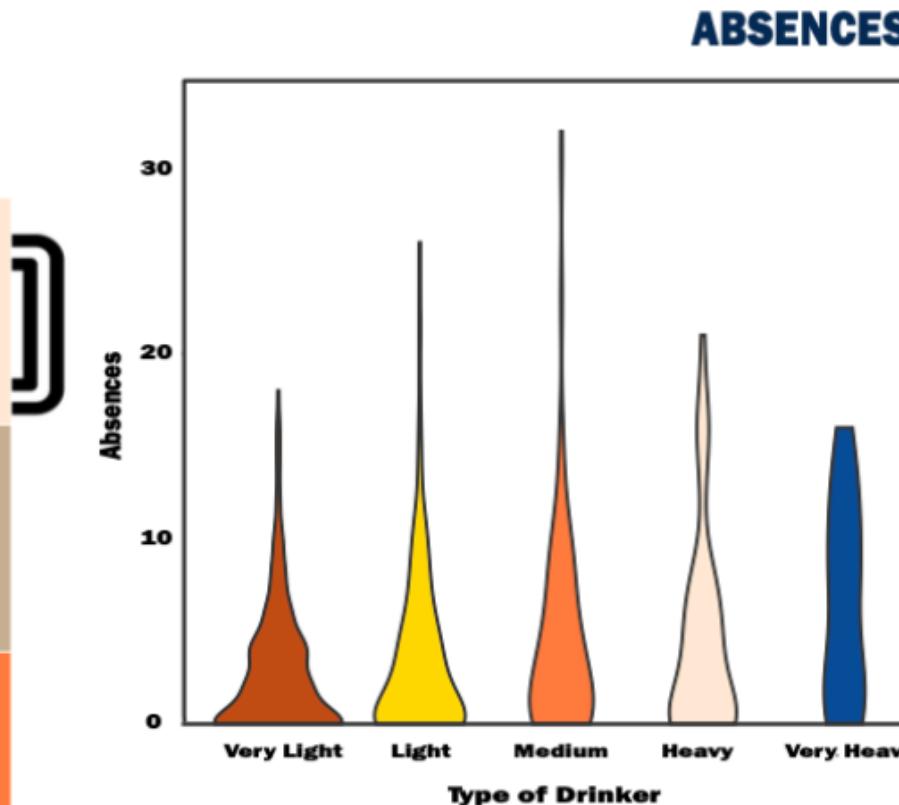


DOES UNDERAGE DRINKING AFFECT A STUDENT'S PERFORMANCE?



The type of drinkers were distributed into categories or levels based on their level of alcohol consumption per week. "Very Heavy" drinkers drank more than 7 times a week. Each type of drinker is drawn to a ranking based on their final semester grade as if they score below average, average, or above average. It's clear that very heavy drinkers perform either average or have a low grade while very light drinkers score above average.

WHAT ARE THEIR BEHAVIORS?



SOURCES:

Student Alcohol Consumption <https://www.kaggle.com/uciml/student-alcohol-consumption>

R PACKAGES:

ggplot2, tidyverse, ggalluvial, dplyr, tidyr
Images: The Noun Project

LEARNING GOALS APPLIED

1. APPLY VISUALIZATION AND PREDICTIVE MODELS TO HELP GENERATE ACTIONABLE INSIGHT

- Visualizations are critical for breaking down large data sets and making data more understandable to the human eye.
- Performed descriptive statistics and transformed them into visualizations to spread a message.

2. COMMUNICATE INSIGHTS GAINED VIA VISUALIZATION AND ANALYTICS TO A BROAD RANGE OF AUDIENCES

- RStudio can create much more descriptive plots than bar graphs, pie charts, and line graphs.
- It is important it is to tell a story from our data and how to communicate it to different audiences.
- Created multiple visuals that any human eye can easily read, while trying to get the story across.
- Once a visualization is made, the story should be able to be told in a few seconds.

3. APPLY ETHICS IN THE DEVELOPMENT, USE AND EVALUATION OF DATA AND PREDICTIVE MODELS

- It's important to have clean, precise, and accurate data to produce an informational graphic that avoids bias and the data from being altered.
- It's not supposed to tell your story, but the data's story.



IST 707: DATA ANALYTICS

COURSE DESCRIPTION

- Led by Professor Yang Yang
- Practiced the fundamentals of data analytics, such as data preparation, concept definition, association rule analytics, classification, clustering, evaluation, and analysis.
- Extracted insights from data using common data analytics approaches while studying the foundations and theories of data analytics methods and how to apply them to solve issues.
- Utilized using open-source software packages in RStudio and Weka.

ABSTRACT

- Student alcohol consumption is a prevalent public health issue, but little is known of its effects on academic achievement.
- Using real-world data obtained from a secondary school in Portugal, our objective was to predict how underage alcohol consumption affects a student's performance.
- Used four different analysis methods: Logistic Regression, Clustering, Classification, and Association Rule Mining to find common variables with students who had high or low alcohol consumption.



1.

REGRESSION ANALYSIS

Explore possible connections or correlations between variables to determine positive or negative relationships.

2.

CLUSTERING

Using k-means clustering we could segment the students based on several metrics. The results of the analysis could be helpful to the school for targeting efforts for student outreach.

METHODS



3.

CLASSIFICATION

Using several prediction methods such as Naive Bayes, Support Vector Machine, Decision Tree, Random Tree to classify accurate relationships.

4.

ASSOCIATION RULE MINING

Finding common variables with students with high and low alcohol consumption metrics.

LOGISTIC REGRESSION

DEPENDENT VARIABLE: ALCOHOL CONSUMPTION (RANGE 2-10)

p ≈ 0

- Increases the more a student goes out (0.59) [1,5]
- Increases if student is male (1.082) [0,1]
- Decreased by quality of family felationships (-0.326) [1,5]
- Increases with absences (0.05) [0,32]
- Health status increases (0.12) [1,5]
- If school choice reason is other increases (0.52) [0,1]
- Increasing study time decreases (0.195) [1,5]
- Family size less than or equal to 3 increases (0.327) [0,1]
- Attending nursery school decreases (-0.3567) [0,1]
- Mothers job in services increases (0.632) [0,1]
- Age increases (0.1) [15,22]

p < 0.05

ASSOCIATION RULES



- Above average alcohol consumption associates with males, high absences, non-parental guardians, and choosing school due to proximity.



- Zero alcohol consumption associates with low going out, wanting to achieve higher education, attending nursery school, and receiving extra educational support.



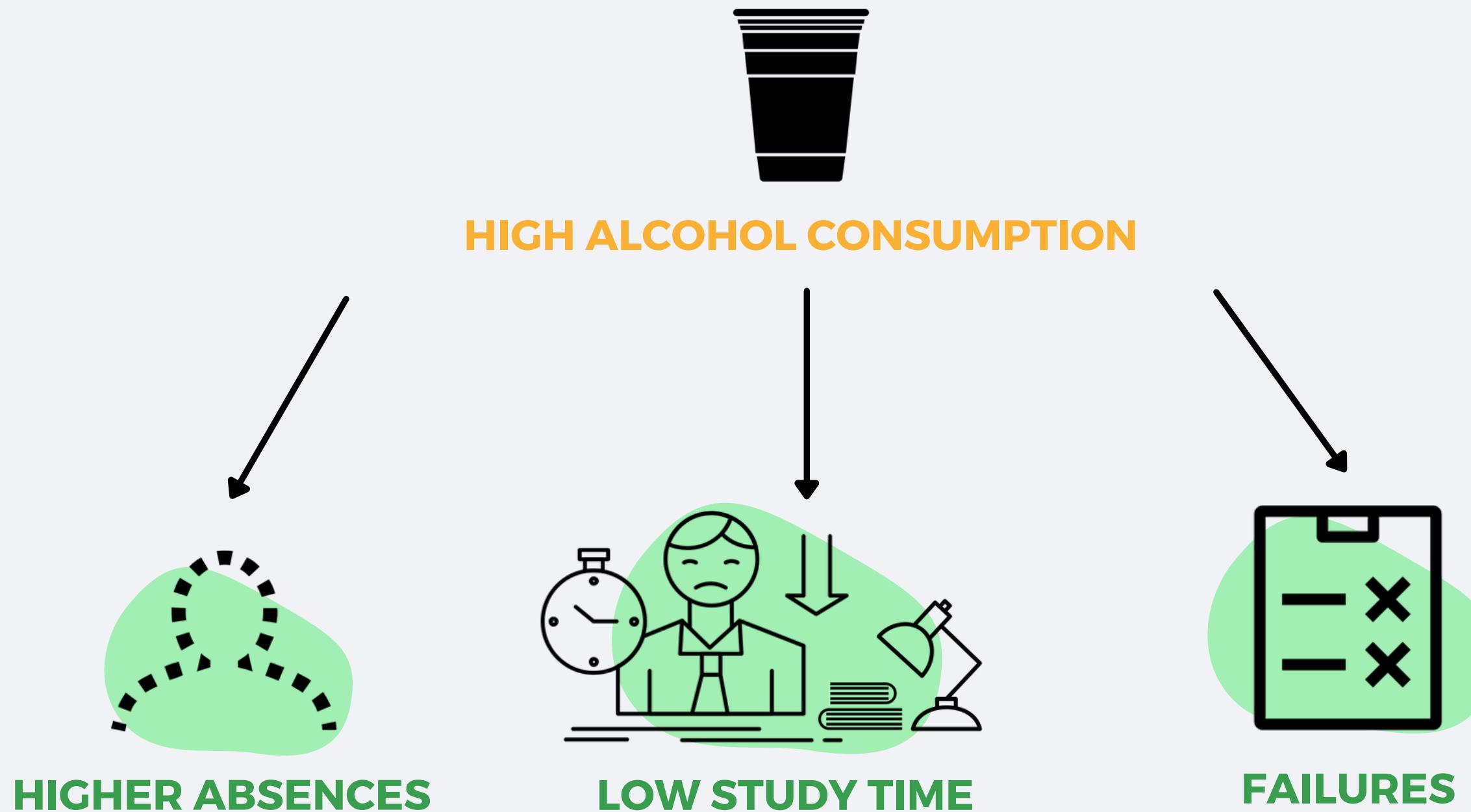
- Low absences associates with low alcohol consumption, no romantic relationship, not paying for extra tutoring, and the mother having less than a primary education.



- High absences associates with at least one class failure, non-parental guardian, having internet access at home, and scoring within 15% of the average.

RESULTS & INSIGHTS

- Underage drinking does affects students, but not dramatically.
- Answered research question, “Can other variables besides alcohol consumption affect student performance?”
 - Identified the significant variables that contributed to higher rates of alcohol consumption.



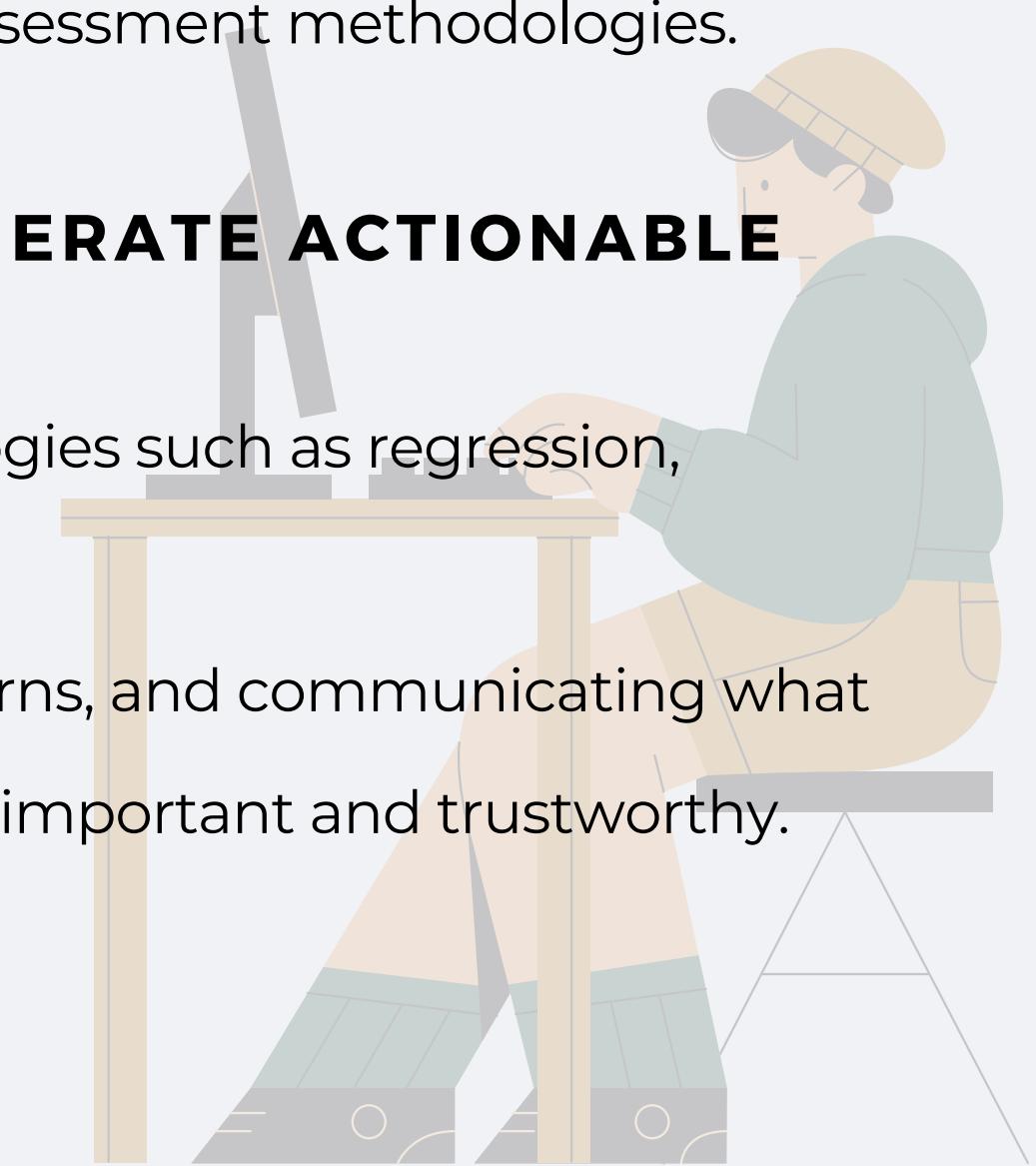
LEARNING GOALS APPLIED

1. CREATE ACTIONABLE INSIGHT ACROSS A RANGE OF CONTEXTS USING DATA AND THE FULL DATA SCIENCE LIFE CYCLE

- Defined, evaluated, and transformed the issues related to data analytics into technological designs and solutions.
- I applied real-world challenges to data analytics principles, algorithms, and assessment methodologies.

2. APPLY VISUALIZATION AND PREDICTIVE MODELS TO HELP GENERATE ACTIONABLE INSIGHT

- To employ data storytelling, I was able to experiment with several methodologies such as regression, classification, and machine learning methods.
- Predictive models are critical for delving into data, discovering relevant patterns, and communicating what patterns have been discovered, how they were discovered, and why they are important and trustworthy.



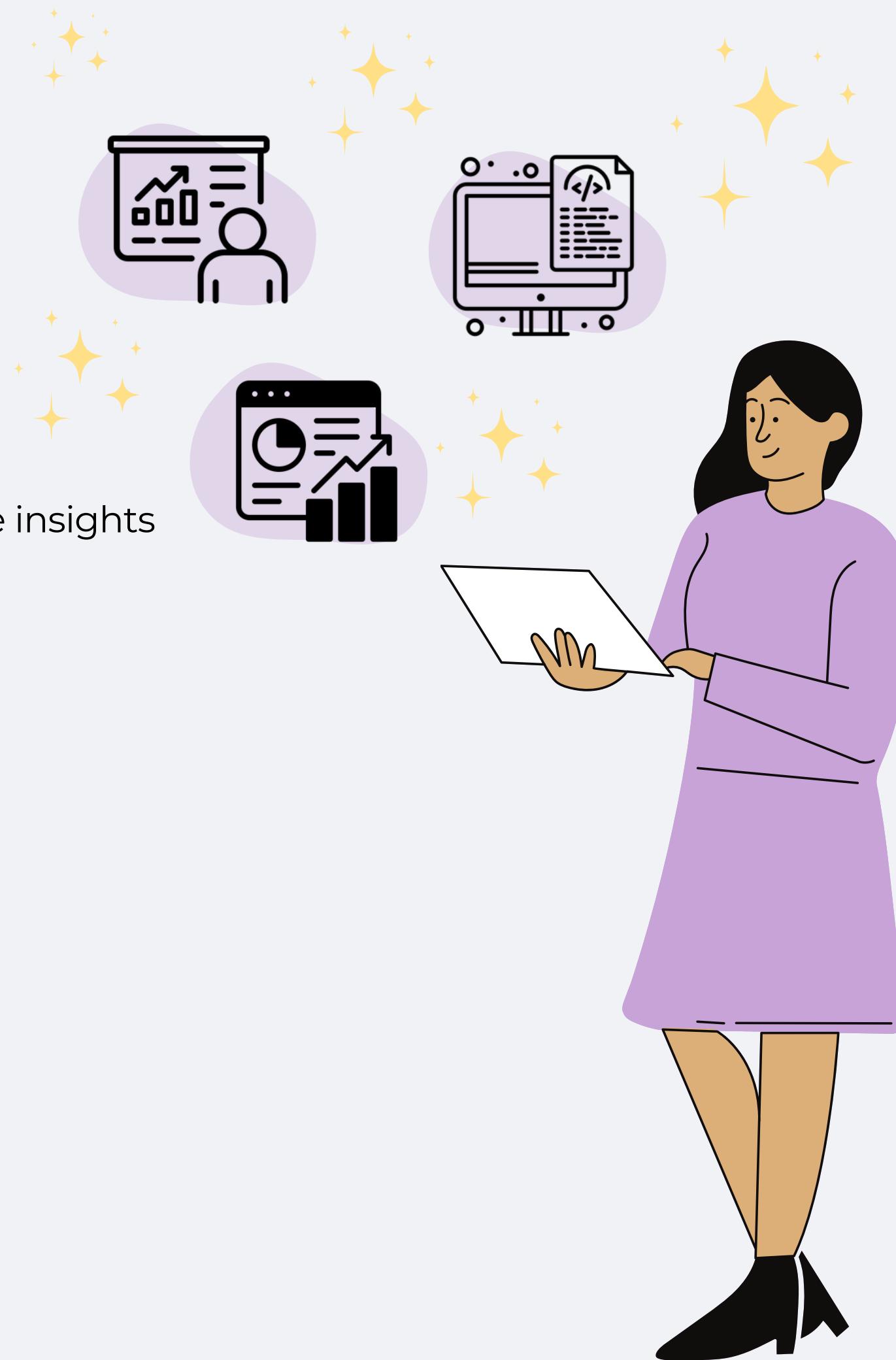
CONCLUSION

TAKEAWAYS

- Data analytics can turn numbers into a story, identify problems, and generate insights
- Essential for powering business decisions and futures
- Many sectors and focus areas under Data Science
- Challenging experience, but extremely rewarding

FUTURE STEPS

- Focus further onto the power of data visualization
- Obtain a career in analytics & visualization
- Be a women in tech!





THANK YOU!
FOREVER ORANGE