



Master of Science

Applied Data Science

Portfolio Milestone
Amanda Norwood

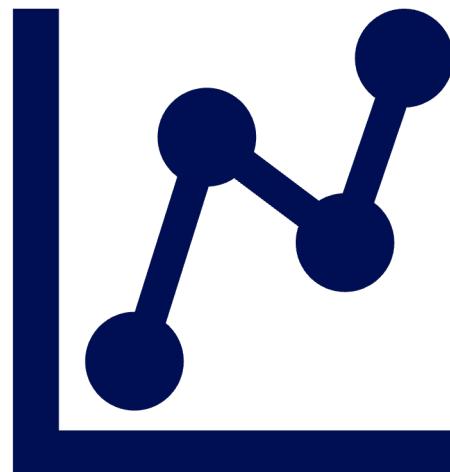
March 2024

503480562



Introduction

Data science is the combination statistics, programming, and advanced analytics to derive actionable insights that's within an organizations data (IBM).





Learning Outcomes

1. Collect, store, and access data by identifying and leveraging applicable technologies.
2. Create actionable insight across a range of contexts (e.g. societal, business, political), 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 (including project sponsors and technical team leads).
6. Apply ethics in the development, use and evaluation of data and predictive models (e.g., fairness, bias, transparency, privacy).

Highlighted Reports & Projects



IST 652
SCRIPTING FOR
DATA ANALYSIS



IST 659
DATABASE
MANAGEMENT



IST 707 APPLIED
MACHINE
LEARNING



IST 718 BIG DATA
ANALYTICS

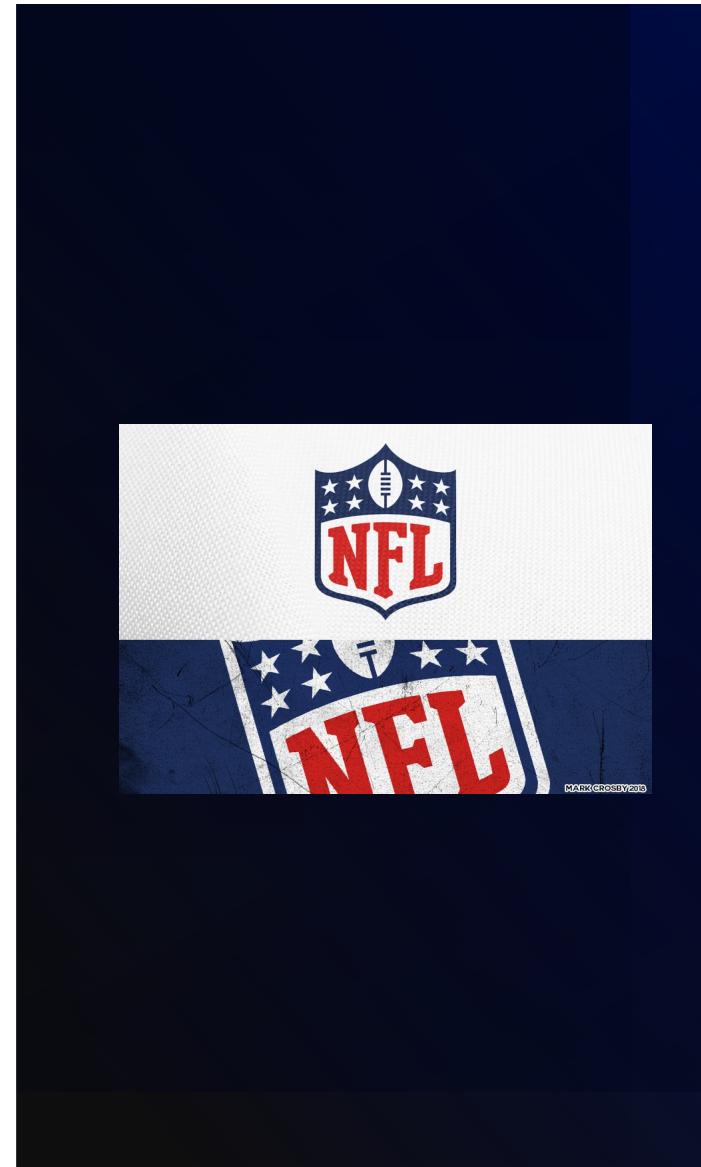


IST 737 VISUAL
ANALYTICS

IST 652 – Scripting for Data Analysis

Project Description

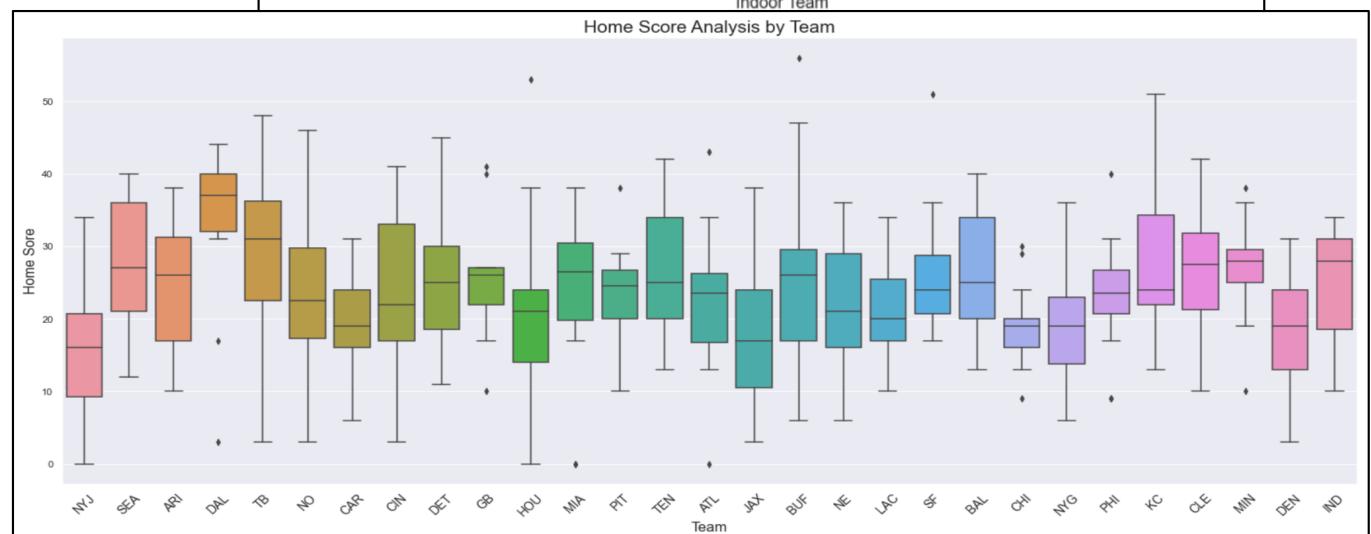
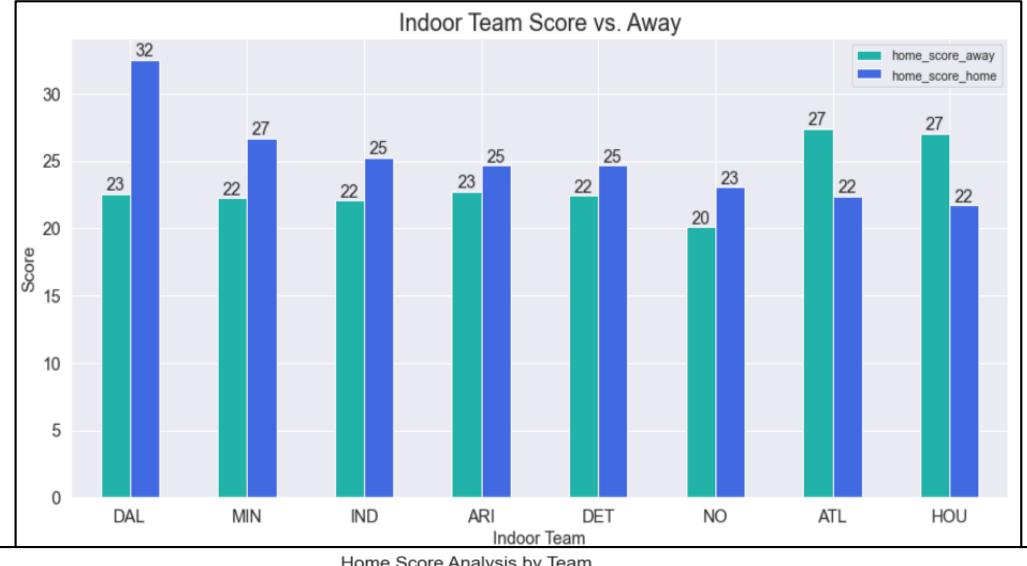
- NFL data analysis from 1997 onwards
- Quarterback-specific data exploration
- Understanding impacts and advantages of different field conditions
- Actionable insights for NFL organization



Individual statistics do not tell the full picture

Data must tie together

Dallas Cowboys perform well at home, average away compared to peers



Learning Outcomes



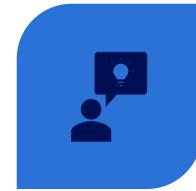
EXPANDED PYTHON PROFICIENCY BEYOND BASIC DATA CLEANING AND CALCULATIONS



DEVELOPED STRATEGIC THINKING IN JOINING AND MERGING MULTIPLE DATASETS



EMPHASIZED THE IMPORTANCE OF CONSISTENT NAMING CONVENTIONS AND THOROUGH DATA CLEANING



GAINED INSIGHTS INTO EXTRACTING ACTIONABLE INSIGHTS FOR IMPACTFUL DECISION-MAKING



ENHANCED UNDERSTANDING OF CONTEXTUAL ANALYSIS AND STORYTELLING WITH DATA



IST 659

Database Management

Project Description

- Objective to manage movie production information and track post-release performances
- Aimed to organize data, reduce clerical errors, and maintain database integrity
- Solution sought to address version control issues faced with spreadsheets
- Mission: Build a database to store information, provide functionality, and ensure data integrity

Examples

ER Diagram:

```

    erDiagram
        movie ||--o{ rating : "fan rating"
        movie ||--o{ rating : "critic rating"
        movie ||--o{ location : "filming"
        movie ||--o{ genre : "is in"
        movie ||--o{ prodCompany : "produced by"
        movie ||--o{ movieCast : "casts in"
        movie ||--o{ person : "has in"
        rating ||--o{ movie : "fan rating"
        rating ||--o{ movie : "critic rating"
        location ||--o{ movie : "filming"
        genre ||--o{ movie : "is in"
        prodCompany ||--o{ movie : "produced by"
        movieCast ||--o{ movie : "casts in"
        person ||--o{ movie : "has in"
        location {
            o--o{ rating : "location"
            o--o{ genre : "location"
            o--o{ prodCompany : "location"
        }
        genre {
            o--o{ movie : "genre"
            o--o{ movieCast : "genre"
        }
        prodCompany {
            o--o{ movie : "prodCompany"
            o--o{ movieCast : "prodCompany"
        }
        movieCast {
            o--o{ person : "person"
            o--o{ role : "role"
        }
        role {
            o--o{ person : "person"
        }
    }
  
```

Table Definitions:

Table	Primary Key	Attributes
Rating	rating_id	fan_rating (R), critic_rating (R)
Location	location_id	location_type (RU), film_rental_rate (R), filming_date_in (R), filming_date_out (R), address (RC)
Genre	genre_id	name (RU)
Prod. Company	prodcompany_id	name (RU)
Movie Cast	moviecast_id	create_date (R), actors (R), director (R)
Role	role_id	role_type (RU)
Persons	person_id	name (RC), gender (R), roles (R)
Film Location	PK, FK1, FK2, FK3	location_id (NOT NULL), movie_id (NOT NULL), locationtype (NOT NULL), date_in (date NOT NULL), date_out (date NOT NULL), location_movieid (NOT NULL)
Genre	PK, FK1	genre_id (NOT NULL), genre_name (NOT NULL), movie_id (NOT NULL)
Prod. Company	PK, FK1	prodcompany_id (NOT NULL), prodcompany_name (NOT NULL), movie_id (NOT NULL)
Movie Cast	PK	moviecast_id (NOT NULL), createdate (datetime)
Person	PK, FK1, UI	person_id (NOT NULL), first_name (NOT NULL), last_name (NOT NULL), gender (char(1) NOT NULL), rate_money (NOT NULL), actor_id (NOT NULL)

BUDGETINFO VIEW:

```

create view budgetinfo as (
    select movietitle, initial_budget, actual_budget,
    (case
        when initial_budget > actual_budget then 'Under Budget'
        when initial_budget < actual_budget then 'Over Budget'
        else 'On Budget Target'
    end) 'Budget_Status',
    (case
        when total_revenue < 0 then 'No Profit'
        when total_revenue > 0 then 'Profit'
        else 'BreakEven'
    end) 'Profit_Status'
    from budget
)
GO
  
```

Results:

moviedb	movietitle	genre	budget	revenue	yearrelease	runtime	prodcompany	director	
1	66	Ghosts in the Wild	Horror	250000.00	NULL	NULL	Olive Butter	2	
1	66	Ghosts in the Wild	Horror	250000.00	582018.00	2022	164	Olive Butter	2



Set Project Budget

Movie Project	▼
Select Director/ Directors	\$. 00
Select Actors/ Support Actors	\$. 00
Select Scene Locations	\$. 00
Select Production Company	\$. 00
Add New Category	
Submit!	

Precision Management Production Location

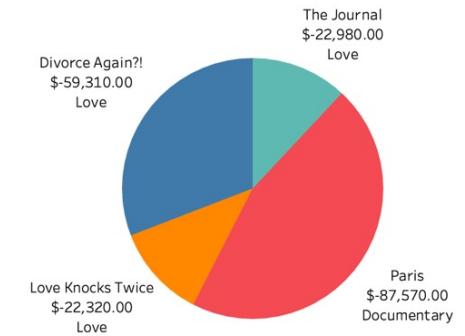
Filters Location Type V

Rental Price Range Min Max

<Location Name>
<Location Type Description>
<Location Rental Range Description>



Movies with No Profit



Linking Tableau and Front-End Development

Learning Outcomes



Appreciation for data engineers' role in database creation and maintenance



Understanding of database structure for efficient query execution



Proficiency in entity relationships and constraints for database integrity



Skill in developing and executing procedures to reduce errors



Knowledge of indexing, denormalization, and query optimization



Understanding of common table expressions and their applications



Recognition of conceptual and logical models' importance



Ability to collect, store, and access data effectively using SQL and Tableau

IST 707 Applied Machine Learning

Project Description

- Bank provided demographic data for customer attrition study
- Analyzed data to uncover reasons for customer churn
- Identified indicators of likely service discontinuation
- Insights aid marketing in personalized recommendations
- Helps refine products/services and enhance satisfaction



Learning Outcomes



Highlighted machine learning importance



Used multiple algorithms, learned about overfitting



Opted for R, found its visualizations appealing



Explored data visualization methods



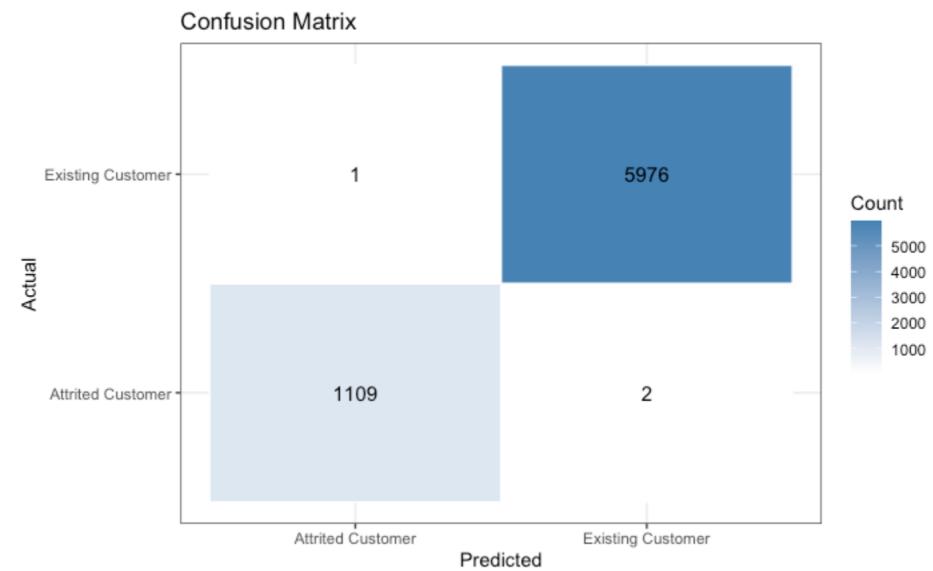
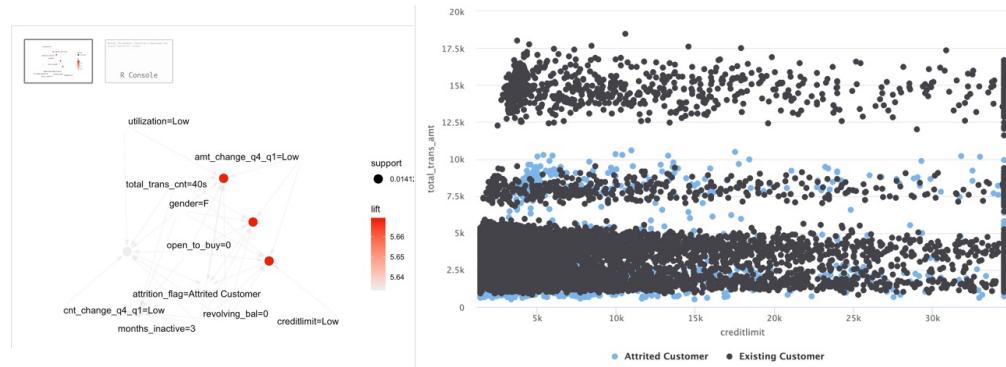
Experimented with machine learning concepts



Importance of meticulous scrutiny for data reliability



Clear definitions crucial when presenting findings



IST 718 BIG DATA ANALYTICS

Los Angeles aims to tackle housing crisis by analyzing affordability

Focus on gender, degree, and profession for analysis

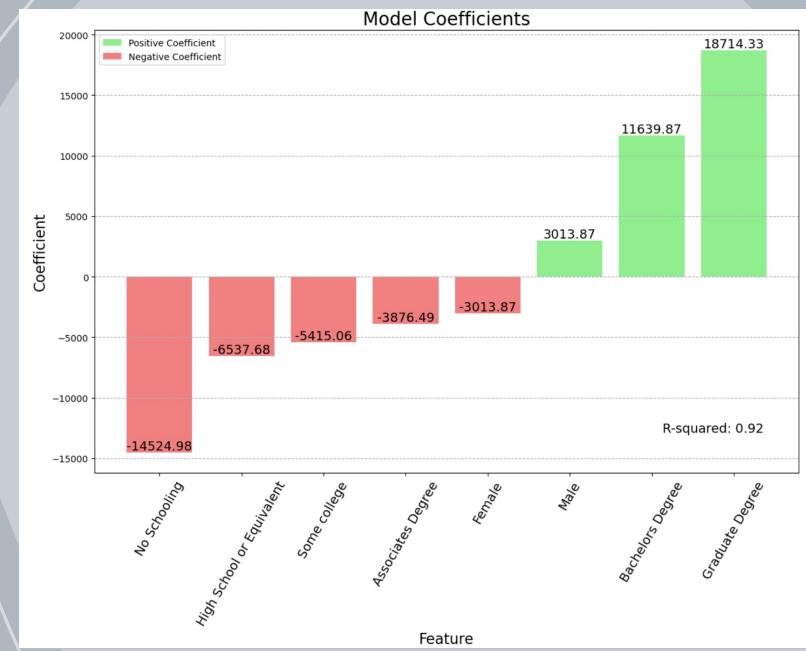
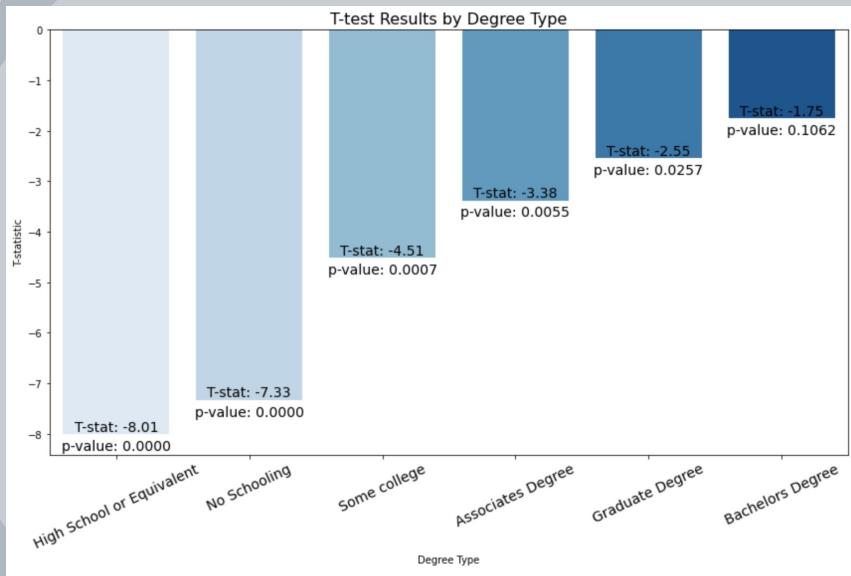
Goal: Assess housing affordability and crime by zip code

Assist potential migrants in researching pay based on profession

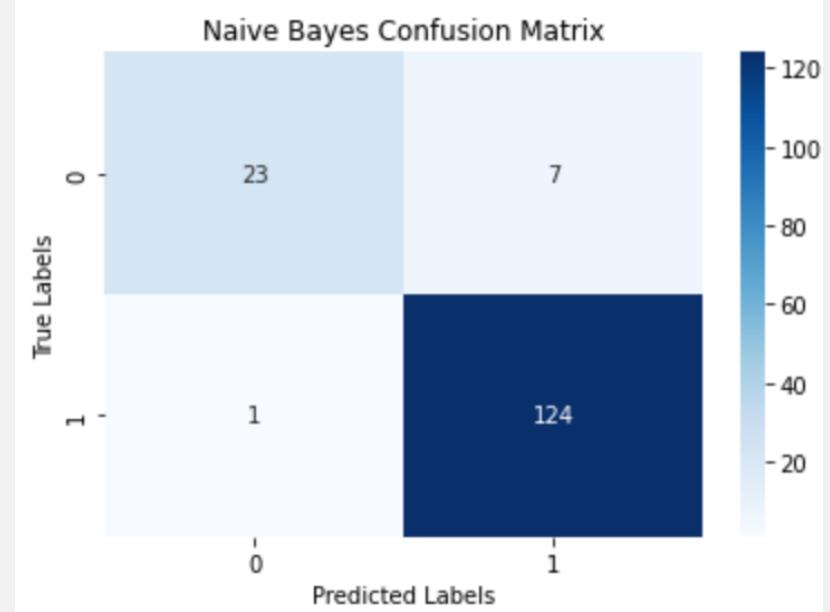
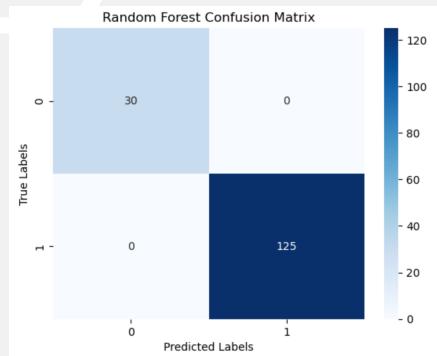
Predict suitable zip codes based on demographic data



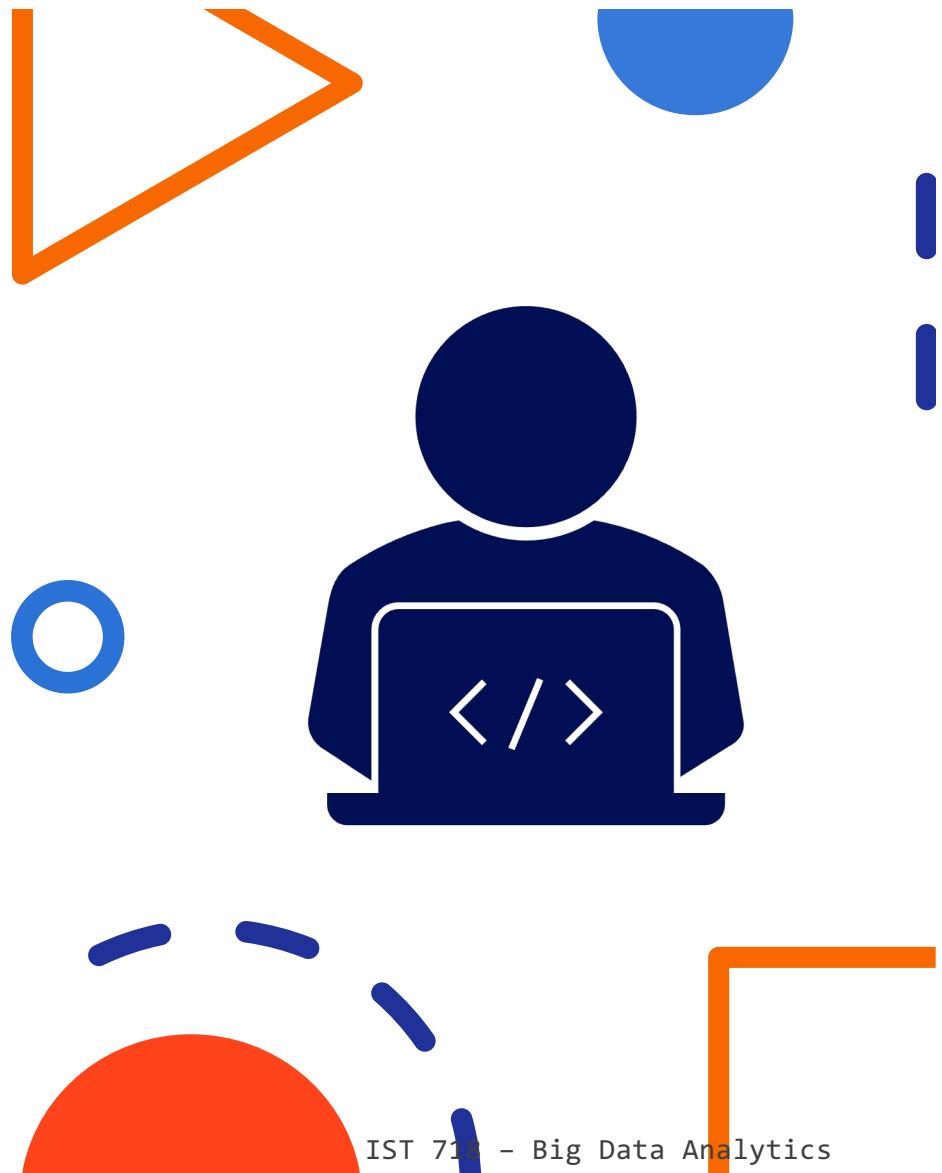
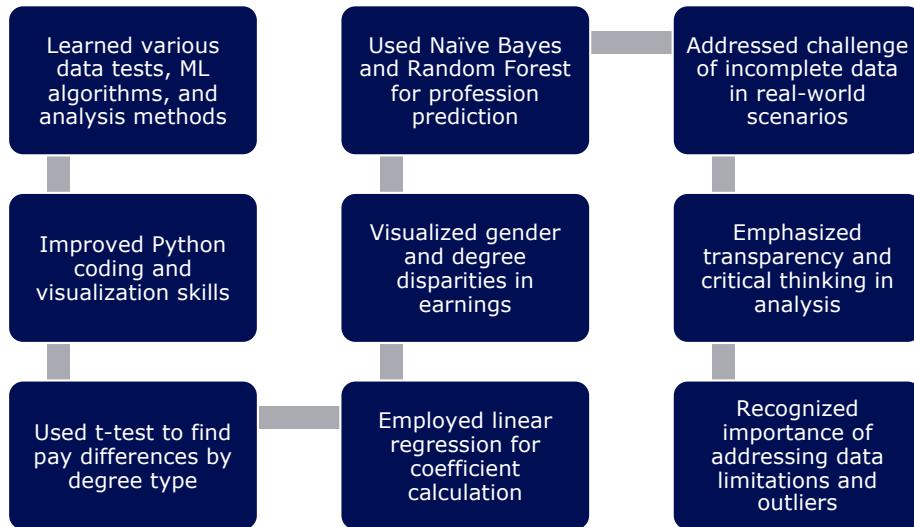
Exploring Gender Pay Disparities Across Degree Types in Los Angeles



Predictive Models in Machine Learning: Naïve Bayes and Random Forests for Professions and Housing Affordability



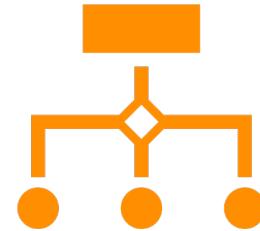
Learning Outcomes



IST 737 Visual Analytics

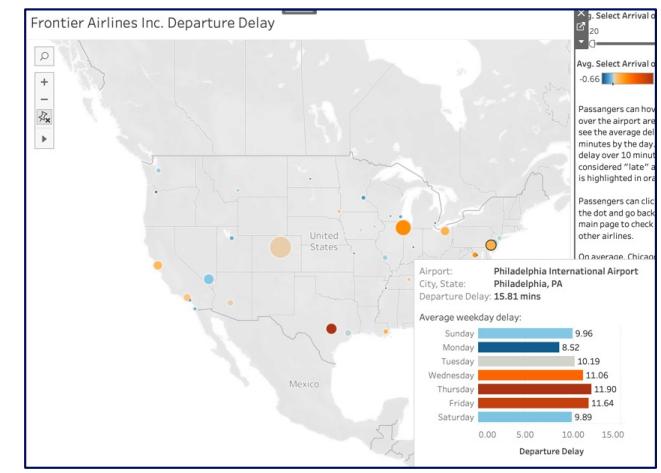
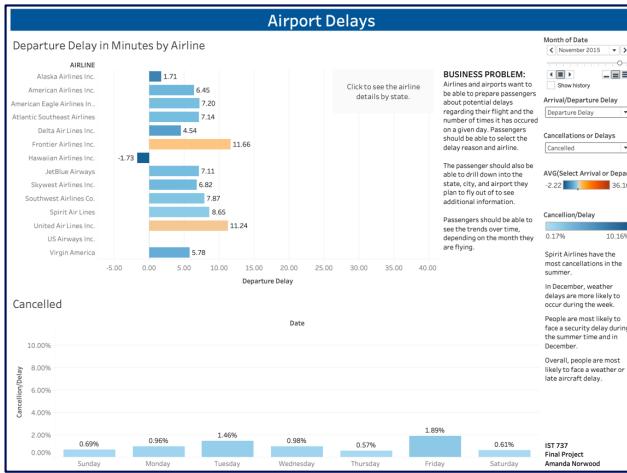


Objective: Develop a system for airports and airlines to track and inform passengers about potential delays or cancellations based on various factors such as airline, airport, day, or state.



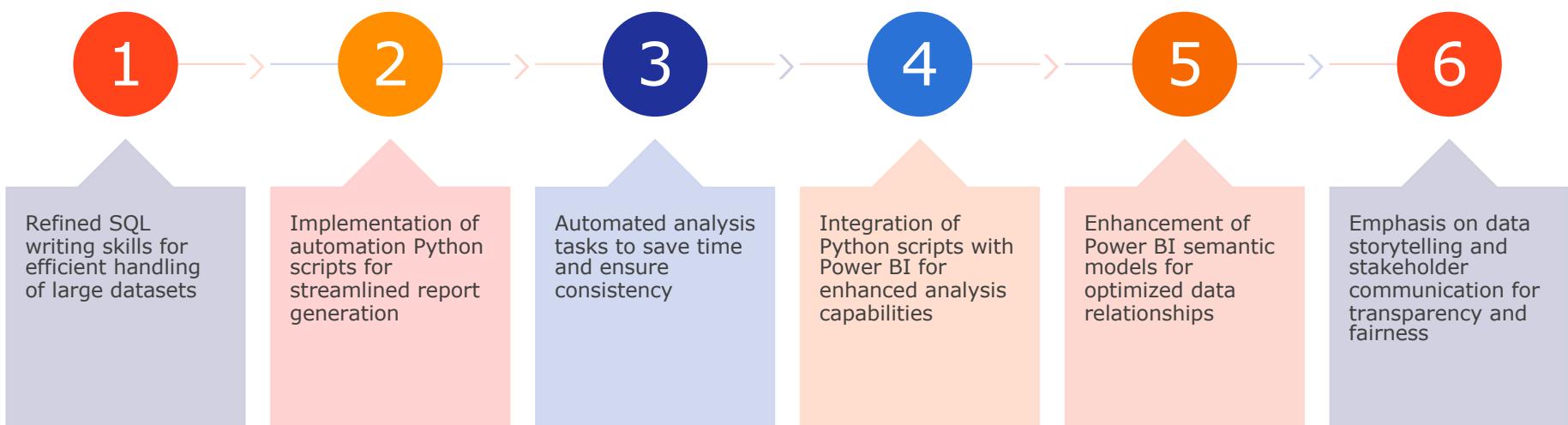
Provide users with both high-level overviews and detailed data by allowing them to drill down into selected information.

Airline Data



Conclusion

Workplace Milestones



Conclusion

Equipped for Real-World Data Challenges



Applied Data Science
program at Syracuse
University's School of
Information Studies



Comprehensive skill set in
data science: collection,
storage, access



Proficiency in modern
technologies: Python, R,
SQL, Power BI, Tableau



Emphasis on OSEMN
method for systematic data
analysis



Diverse project experience
across database
management, analytics,
and machine learning



Interdisciplinary approach
and communication skills
emphasized



Master of Science

Applied Data Science

Amanda Norwood

