PIC Math: Industry Projects for Ownership, Engagement, and Proficiency

Vinodh Chellamuthu
Dixie State University
St. George, UT

Jennifer Travis
Lone Star College-North Harris
Houston, TX

November 13, 2020

Who We Are

Jennifer Travis

- MS in Mathematics, PhD in Mathematics Education
- Lone Star College-North Harris - a large community college

Vinodh Chellamuthu

- MS in Applied Mathematics, PhD in Mathematics
- Mentoring undergraduates in research since 2016
- Dixie State University
  - Public four-year university in St. George, UT
  - Associate-granting institution
  - Open access; lots of first-generation college students

What is PIC Math?

PICMath

Preparation for Industrial Careers in Mathematical Sciences

Each PIC Math faculty member teaches a class in which students work in teams to solve a a real-world problem provided by a business, industry, non-profit, or government partner.

- Emphasis on statistics and data analysis, with data provided by the partner organization.
- Builds skills needed to help students stand out in the job market and kickstart a successful career.
- Training and support provided by the MAA PIC Math Program, which is funded by the National Science Foundation (NSF grant DMS-1722275) and National Security Agency.

Fast Facts on PIC Math Program

- Started in 2014
- 5 cohorts of PIC Math faculty so far
- Over 180 total PIC Math faculty
- Over 150 institutions
- 6 community colleges

For more information:
`https://www.maa.org/programs-and-communities/professional-development/pic-math`

Why is PIC Math Valuable for Students?

- Allow students to get hands-on experience and introduce them to more career possibilities.
- Provide an opportunity for students to use the skills they learn in classes to solve messy and complex real-world problems.
- Develop appreciation for the value of mathematics and data analysis in solving practical problems.
- Learn professional skills (writing, presentation, teamwork).

More Information on PIC Math

For more on WHY the class is valuable:

PIC Math and the Best Jobs in the 21st Century
by Michael Dorff and Suzanne Weekes
`https://youtu.be/B5M4ByxtEdk`

For more on HOW the class works:

Equipping Students for STEM Careers through Industrial Projects: The MAA PIC Math Program
by Vinodh Chellamuthu and Jennifer Travis
`https://youtu.be/aKC-HfEmNtM`

Structure of the Class

Vinodh

- Taught Spring and Fall of 2019 and 2020
- MATH 4800 (Industrial Careers in Mathematics)
- 3 credit hours
- Letter grade
- Counted toward degree plan for several STEM majors
- Served as an "applied learning" project class for a new degree program in Applied and Computational Mathematics

Jennifer

- Taught Spring 2020
- Honors Seminar
- 4 Honors credits
- Pass/No Pass
- Met TTh 1.5 hours
- Did not count toward a degree plan

The Students

Vinodh

- 30 students across 4 semesters
- 7 math majors, 8 engineering, 3 computer science, 1 finance
- Majority first-generation college students
- Some in Multivariable Calculus, some in Linear Algebra, some in introductory programming classes

Jennifer

- 11 students in one class so far
- 9 first-generation college students
- 1 math major, others variety of STEM
- 4 in their first year of college
- Half in Calculus, others in College Algebra, Trigonometry, Precalculus
- Almost all had zero programming experience
- Half already in Honors, half joined Honors to take class

Problems Tackled by our Students - Jennifer

Houston Astros:
Provide the Astros with recommendations on what types of fan promotions are most effective at increasing home game attendance.

Problems Tackled by our Students - Jennifer

Crown Beverage Packaging:
Analyze spoilage data and provide company with recommendations for how they can improve or better understand can spoilage at the Conroe plant.

Our Experiences - Jennifer

Computing Tools

Data cleaning and organizing:

- Excel

Analysis and plots:

- Python/Jupyter Notebook (installed on hard drive)
  `www.anaconda.com`
- Google Colab (cloud-based equivalent)
  `https://colab.research.google.com`

Reports and slides:

- LaTeX
- Overleaf cloud platform
  `www.overleaf.com`

Converting data to cans:

- Production data was in various units
- Examples
  - 21 Layers = 1 Pallet
  - 1 Layer = 389
  - 1 Pallet = 8169 cans
  - 1 Tote = 584 cans (later changed to 750)
  - 1 Dumpster = 2723 cans
  - 1 Xsmall dumpster = 2156 cans
  - 1 Large Dumpster = 4300 cans
  - 1 Small Dumpster = 2425 cans

Data Organizing with Excel

Jennifer's Houston Astros Students

- Resolved formatting issues
  (e.g., 11-8 Win-Loss record appeared as 8-Nov)
- Normalized Attendance: divided by Stadium Capacity
- Qualitative Data Analysis: Categorizing the Promotions

AU	AV	BE	BF	BL	BN
Giveaway Item	**Second Item**	**Event**	**Second Event**	**Notes**	**Decile Rank**
Jersey		Pride Night		Maryland Flag Jersey (15 and over) / Maryland Day	Decile 10
		Kids Run Bases		Kids Run the Bases	Decile 9
					Decile 2
					Decile 1
T-shirt				T-shirt (all fans)	Decile 7
		Fireworks		Friday Night Fireworks	Decile 9
Figurine				Trey Mancini (15 and over)	Decile 3
Figurine		Kids Run Bases		Garden gnome (15 and over) / Kids Run the Bases	Decile 1
		Fireworks		Friday Night Fireworks	Decile 1
Cards				Topps card pack (all fans)	Decile 1
					Decile 1
Home Item		Kids Run Bases		Alarm clock (all fans 14 and under) / Kids Run the Bases	Decile 2
Bobblehead		GT Night		Game of Thrones Night w/ bobblehead (theme ticket)	Decile 4
T-shirt				T-shirt (all fans)	Decile 10
		Fireworks		Friday Night Fireworks	Decile 6
					Decile 2

Analysis and Plots with Jupyter Notebook and Colab

Jennifer's Students

jupyter Interactive-Plot-and-Crosstables_2020-07-04-a_JT Last Checkpoint: 07/05/2020 (autosaved) Logout

File Edit View Insert Cell Kernel Widgets Help Not Trusted Python 3

Run Code

```
ALEast = ['BOS','NYY','TBR','TOR','BAL']
ALCentral = ['CLE','MIN','DET','CHW','KCR']
ALWest = ['HOU','OAK','SEA','LAA','TEX']

highAttendance = ['BOS','CHC','NYY','HOU']
lowAttendance = ['HOU','MIA','OAK','PIT']
```

In [229]:
```
#dfNLEast = df0[df0.Team.isin(['ATL','WSN','PHI','NYM','MIA'])]  #Another way to do it.
dfNLEast = df0[df0.Team.isin(NLEast)]
dfNLCentral = df0[df0.Team.isin(NLCentral)]
dfNLWest = df0[df0.Team.isin(NLWest)]

dfALEast = df0[df0.Team.isin(ALEast)]
dfALCentral = df0[df0.Team.isin(ALCentral)]
dfALWest = df0[df0.Team.isin(ALWest)]

dfLowAttendance = df0[df0.Team.isin(lowAttendance)]
dfHighAttendance = df0[df0.Team.isin(highAttendance)]


#dfNLCentral  #I reran this 6 times with a different division here. Scrollled through df to see if it looked right.
dfHighAttendance #Also checked it for low attended teams and high attended teams.
```

Out[229]:

	TeamGameID	DataID	Year	Team	Game	Day	Month Date	Opponent	W-L	W	L	Wpct	Lpct	GB	Win	Loss	Save	Tin
243	BOS7	243	2018	BOS	7	Thursday	Apr 5	TBR	1-Jun	6	1	85.71	14.29	2.0	Poyner	Kittredge	NaN	3:5
244	BOS8	244	2018	BOS	8	Saturday	Apr 7	TBR	1-Jul	7	1	87.50	12.50	2.5	Porcello	Faria	NaN	2:4
245	BOS9	245	2018	BOS	9	Sunday	Apr 8	TBR	1-Aug	8	1	88.89	11.11	2.5	Smith	Colome	Kimbrel	3:5

Python Packages

Jennifer's Students

- Pandas - working with data in dataframes
- Seaborn - static plots
  - Simple coding for plots
  - Built-in datasets for practice and examples
- Altair - interactive plots
  - `http://nhmath.lonestar.edu/Faculty/TravisJ/PICMath/ALWest.html`
  - `http://nhmath.lonestar.edu/Faculty/TravisJ/PICMath/cameraJuly2.html`

Identifying Highly-Attended Games

The Formula

To test for outliers, we used the robust z-score (Iglewicz & Hoaglin, 1993):

$$Z = \frac{0.6745(x_i - x)}{MAD}$$

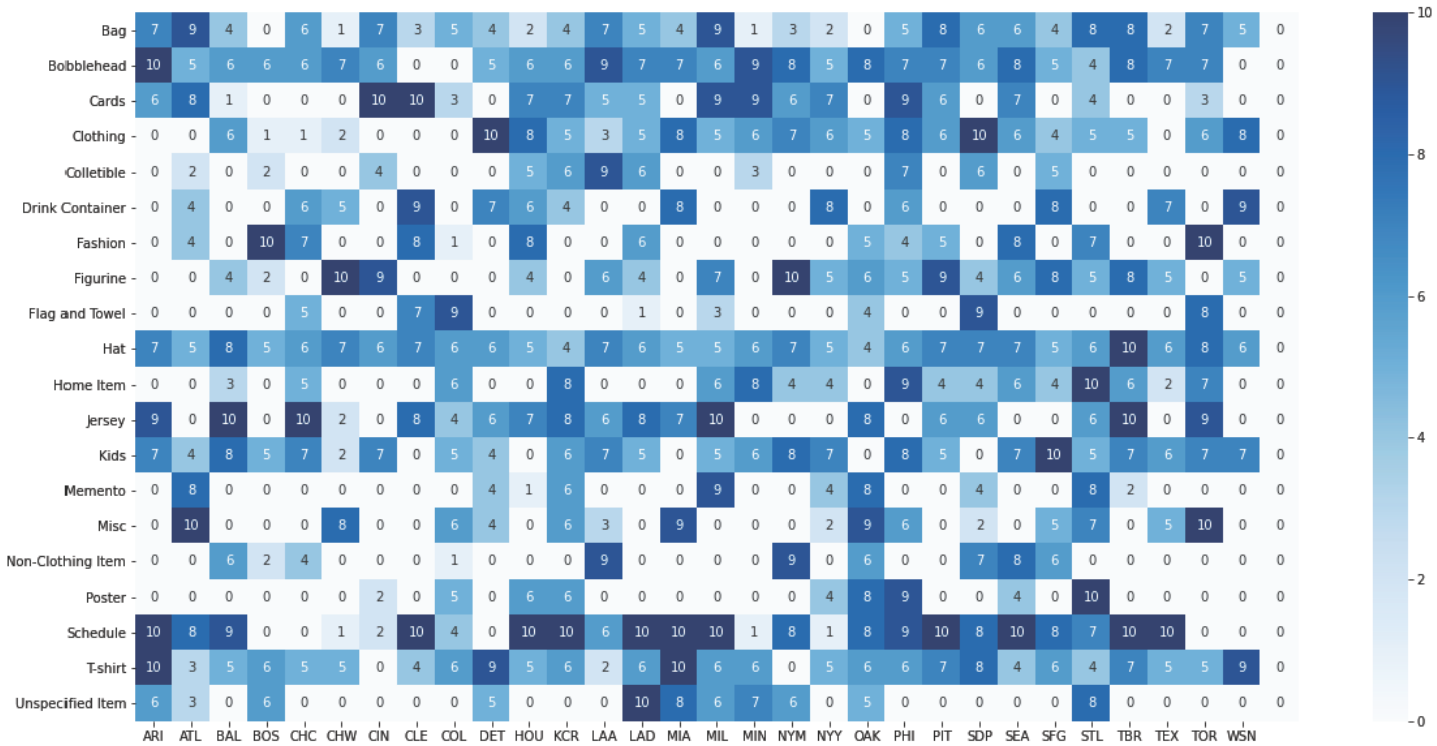
MAD= Median Absolute Deviation

Why We Chose It

It relies on the median instead of the mean, so it isn't as affected by outliers as the traditional z-score.

Outlier Threshold

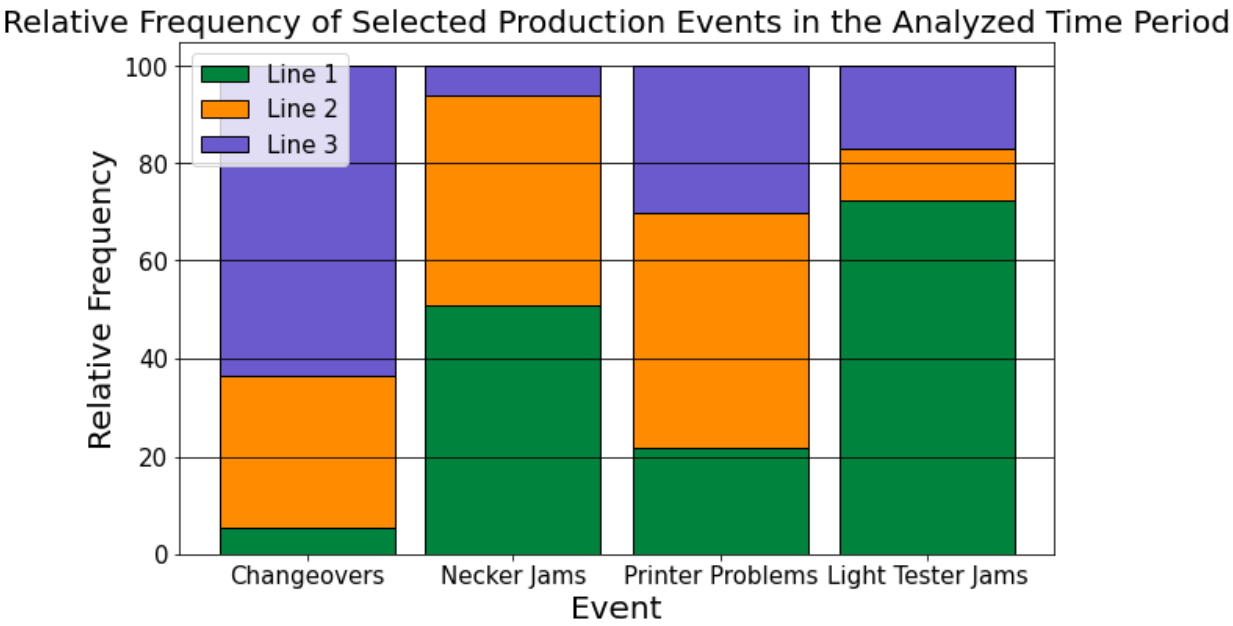
Games with z-score of at least 2.0 were considered outliers.


Average Decile of Giveaways by Team
Bag
Bobblehead
Cards
Clothing
Colletible
Drink Container
Fashion
Figurine
Flag and Towel
Hat
Home Item
Jersey
Kids
Memento
Misc
Non-Clothing Item
Poster
Schedule
T-shirt
Unspecified Item
ARI ATL BAL BOS CHC CHW CIN CLE COL DET HOU KCR LAA LAD MIA MIL MIN NYM NYY OAK PHI PIT SDP SEA SFG STL TBR TEX TOR WSN
10
8
6
4
2
0

Some Plots

Jennifer's Students

Frequencies of Common Production Events

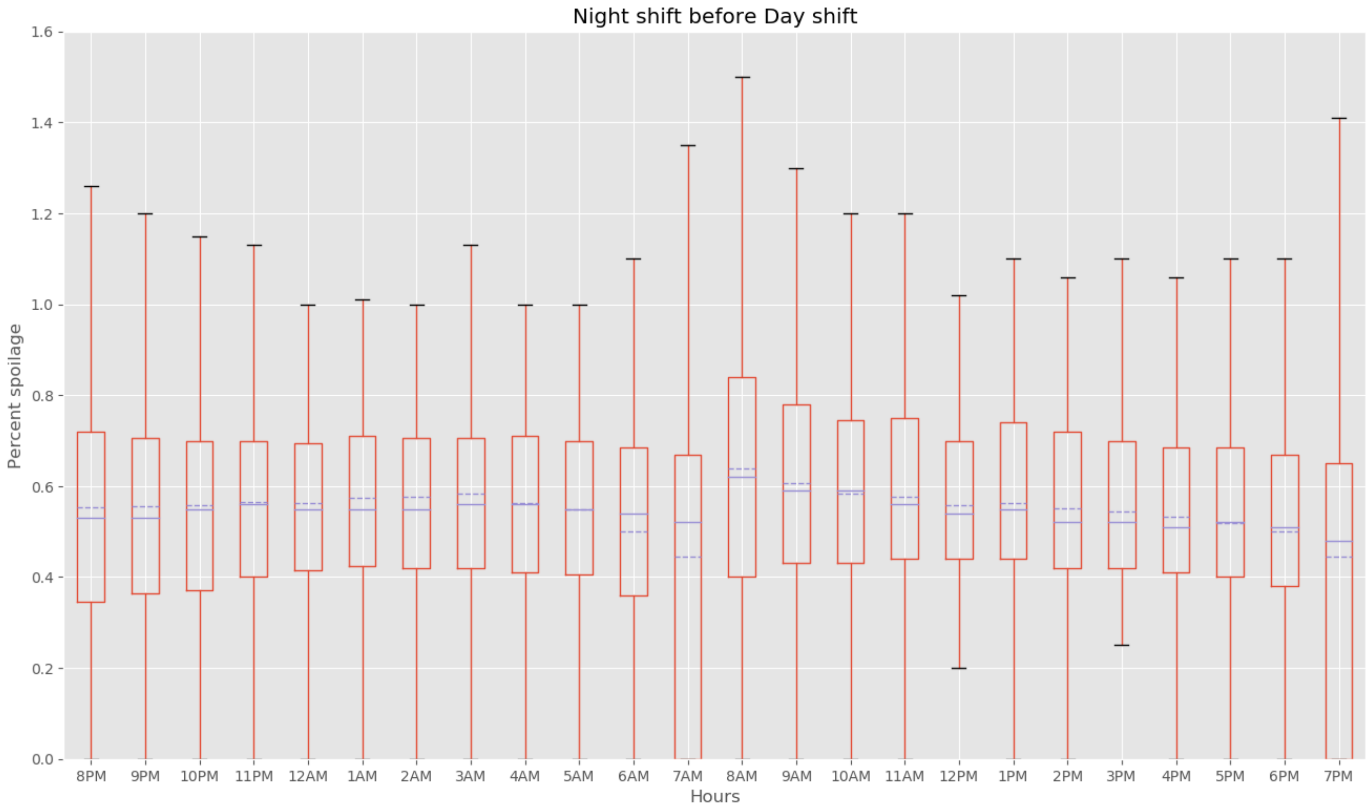


Some Plots

Jennifer's Students

Analysis

Box Plot showing the range of percentages throughout 24 hours at each hour for Line 2 for the whole data set



LaTeX/Overleaf for Slides and Reports

- Overleaf: Cloud platform for compiling LaTeX code.
- Example: `https://www.overleaf.com/read/vxywktbgdbqt`
- Free version worked well.
- Collaborate on shared documents.
- Chat function for late-night group editing sessions.

Trail Activity Prediction Model for Zion National Park

Heather Smith | Douglas Baer | Nicholas Warner

Dixie State University
St. George, UT

Faculty Mentor: Dr. Vinodh Chellamuthu
Industry Liaison: Jason Pitts, Program Director
Industry Partner: Park Data Project

This work is part of the PIC Math Program

DSU DIXIE STATE UNIVERSITY ST. GEORGE, UTAH

PICMath

MAA MATHEMATICAL ASSOCIATION OF AMERICA

2

Problems Tackled by our Students - Vinodh

Parks Data

Background

- Thousands of people come to Zion National Park (ZNP) every day

Park administration, employees, and visitors do not know how busy trails are going to be at each hour of the day

- Trail activity predictions can help park administration:
  - Plan and allocate resources
  - Protect visitors, the natural beauty, and wildlife of the park
  - Improve visitor experience by providing trail activity insights

DSU DIXIE STATE UNIVERSITY ST. GEORGE, UTAH

Problems Tackled by our Students - Vinodh

Parks Data

Datasets

Mar 2015-Feb 2020, Hourly Measure of Activity (HMOA) data for five different trails:

- Angel's Landing
- Riverside Walk
- Watchman
- Pa'rus
- Kayenta

May 14, 2019 Angel's Landing	
12:00 AM	0
1:00 AM	0
2:00 AM	0
3:00 AM	0
4:00 AM	0
5:00 AM	10
6:00 AM	130
7:00 AM	230
8:00 AM	220
9:00 AM	260
10:00 AM	220
11:00 AM	220
12:00 PM	220
1:00 PM	200
2:00 PM	160
3:00 PM	100
4:00 PM	120
5:00 PM	60
6:00 PM	10
7:00 PM	0
8:00 PM	0
9:00 PM	0
10:00 PM	0
11:00 PM	0
Total	2160

Daily Measure of Activity (DMOA)

DSU DIXIE STATE UNIVERSITY ST. GEORGE, UTAH

Problems Tackled by our Students - Vinodh

Parks Data

Sliding Window Algorithm

- *Weather Forecasting Using Sliding Window Algorithm* by Piyush Kapoor and Sarabjeet Singh Bedi (2013)
- Predicts humidity, rainfall, and high and low temperatures by the same method
  - Behavior tomorrow will be similar to today, the current week, and some week around the same time last year
  - Use Euclidean distance to find the most similar week last year
  - Use the average of differences between days to predict for tomorrow

DSU DIXIE STATE UNIVERSITY ST. GEORGE, UTAH

Problems Tackled by our Students - Vinodh

Parks Data

Sliding Window

- *Weather Forecasting Using Sliding Window Algorithm* by Piyush Kapoor and Sarabjeet Singh Bedi (2013)

$$ED_k = \frac{\sqrt{\sum_{i=1}^{n}(q_i - p_i)^2}}{n}$$

Predict HMOA for May 14, 2019

2019

Tue May 7	Wed May 8	Thu May 9	Fri May 10	Sat May 11	Sun May 12	Mon May 13	Sun May 14
1650	2170	2180	1760	3010	2240	2100	?

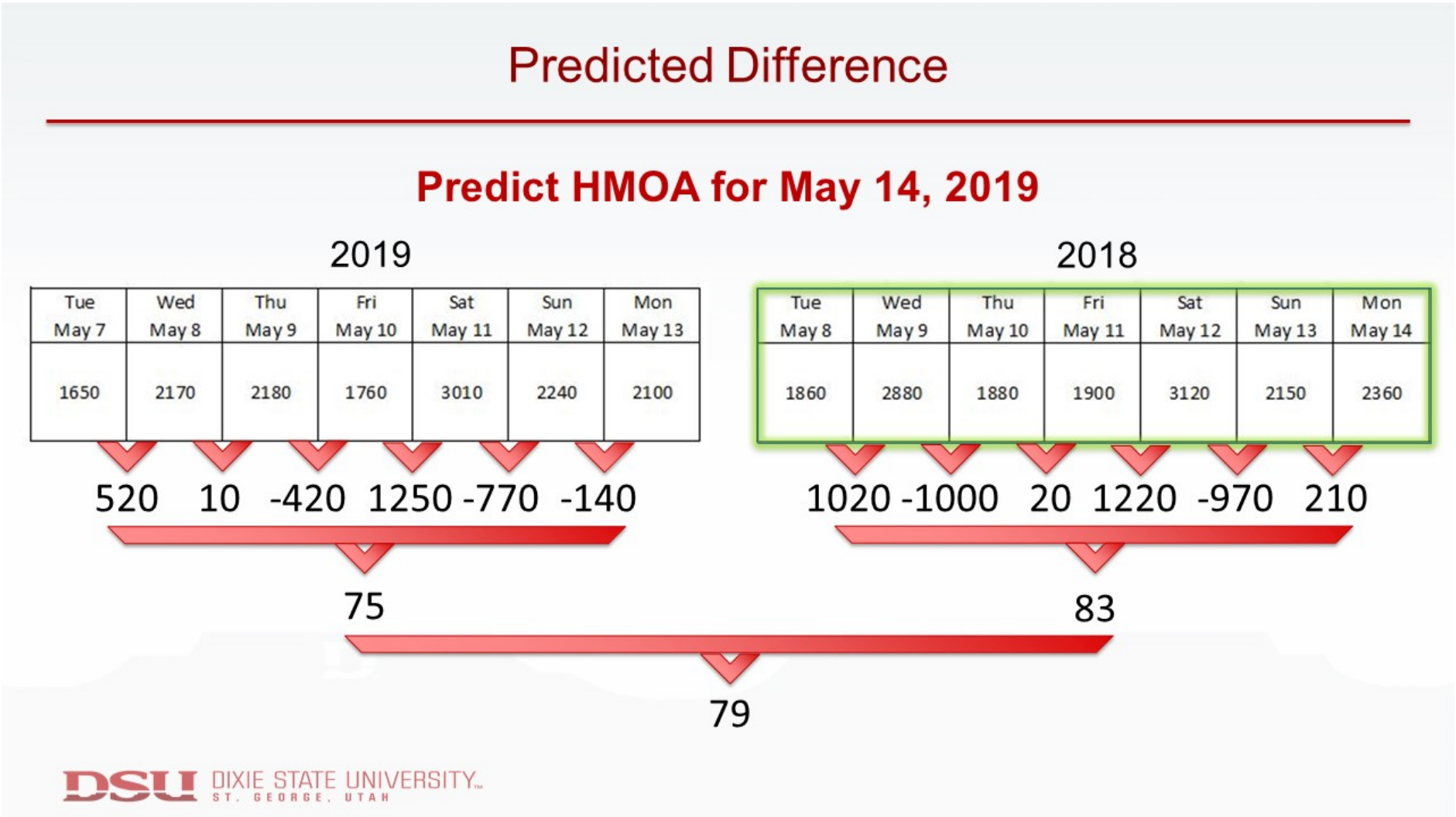
2018

Mon May 7	Tue May 8	Wed May 9	Thu May 10	Fri May 11	Sat May 12	Sun May 13	Mon May 14	Tue May 15	Wed May 16	Thu May 17	Fri May 18	Sat May 19
2130	1860	2880	1880	1900	3120	2150	2360	2400	2190	2440	2680	2990

DSU DIXIE STATE UNIVERSITY ST. GEORGE, UTAH

Problems Tackled by our Students - Vinodh

Parks Data

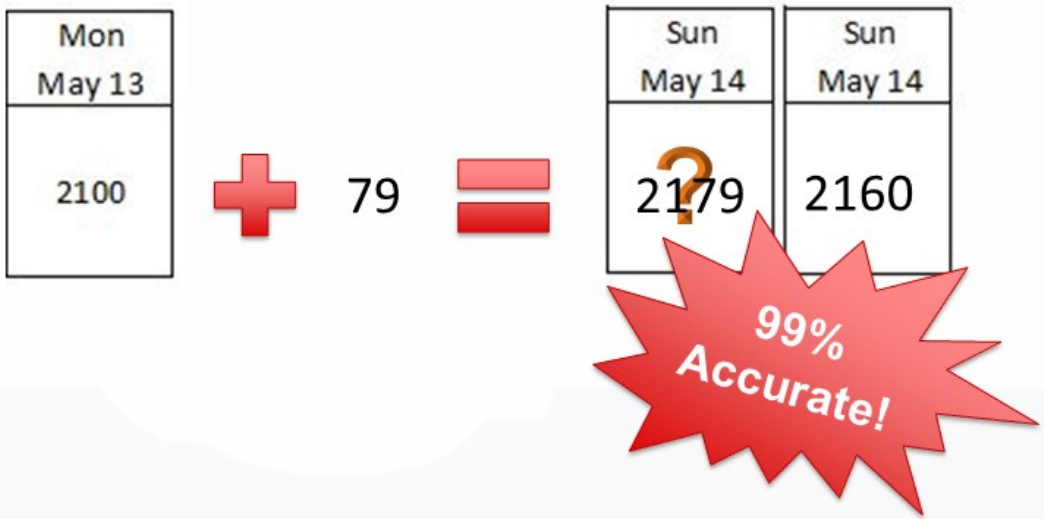


Problems Tackled by our Students - Vinodh

Parks Data

Output

Predict HMOA for May 14, 2019



DSU DIXIE STATE UNIVERSITY ST. GEORGE, UTAH

Hourly Prediction Model

Utilize distribution patterns to make hourly predictions:

1. Take the proportion of each hour:

$$\textbf{Hour Proportion} = \frac{\textbf{Hour HMOA}}{\textbf{DMOA}}$$

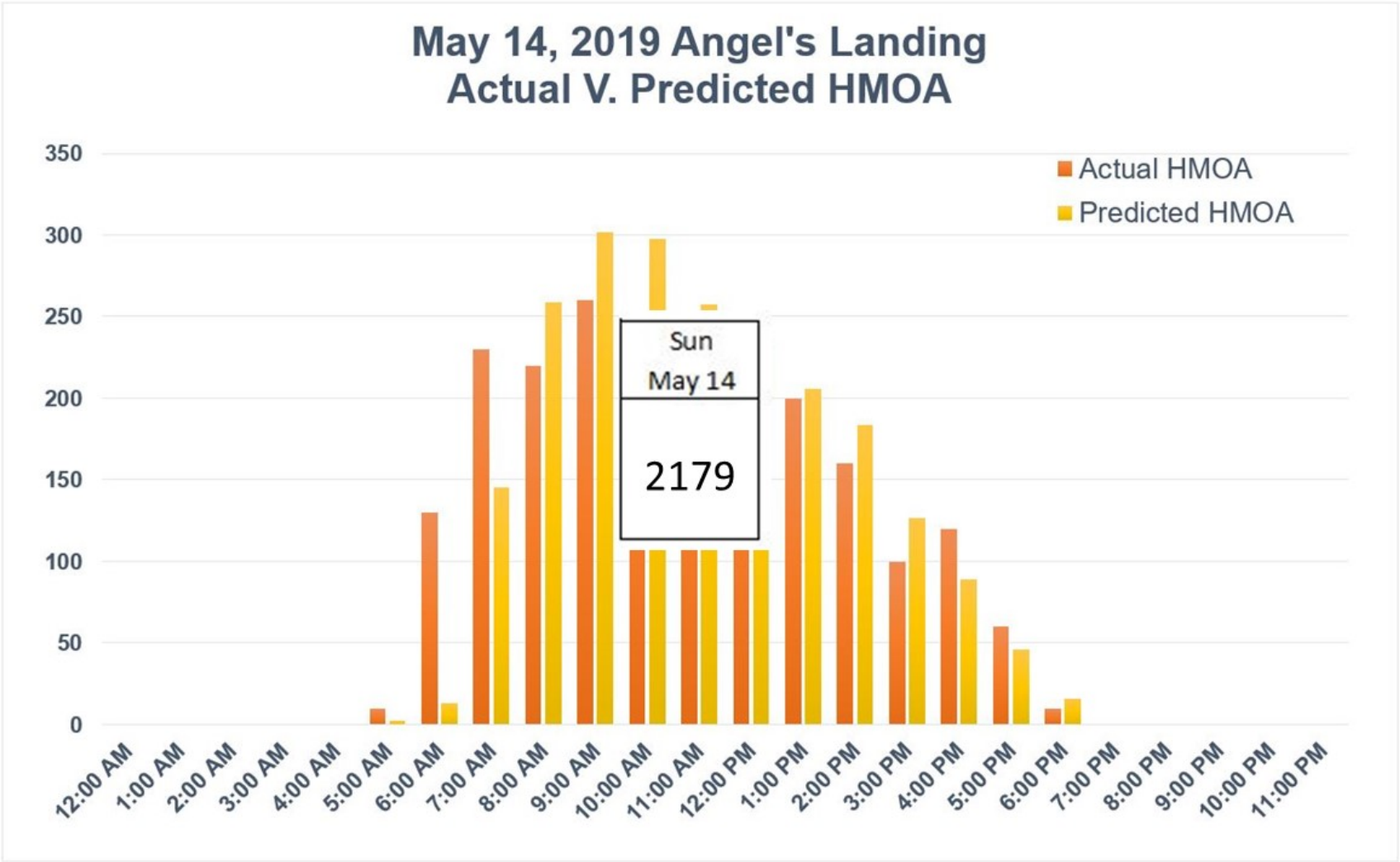
2. Average the proportion of each hour of the previous 31 days (AHP)
3. Use the average of each hour to divide the DMOA prediction into hourly predictions

$$\textbf{Hour Prediction} = \textbf{Daily MOA} \cdot \textbf{AHP}$$

DSU DIXIE STATE UNIVERSITY ST. GEORGE, UTAH

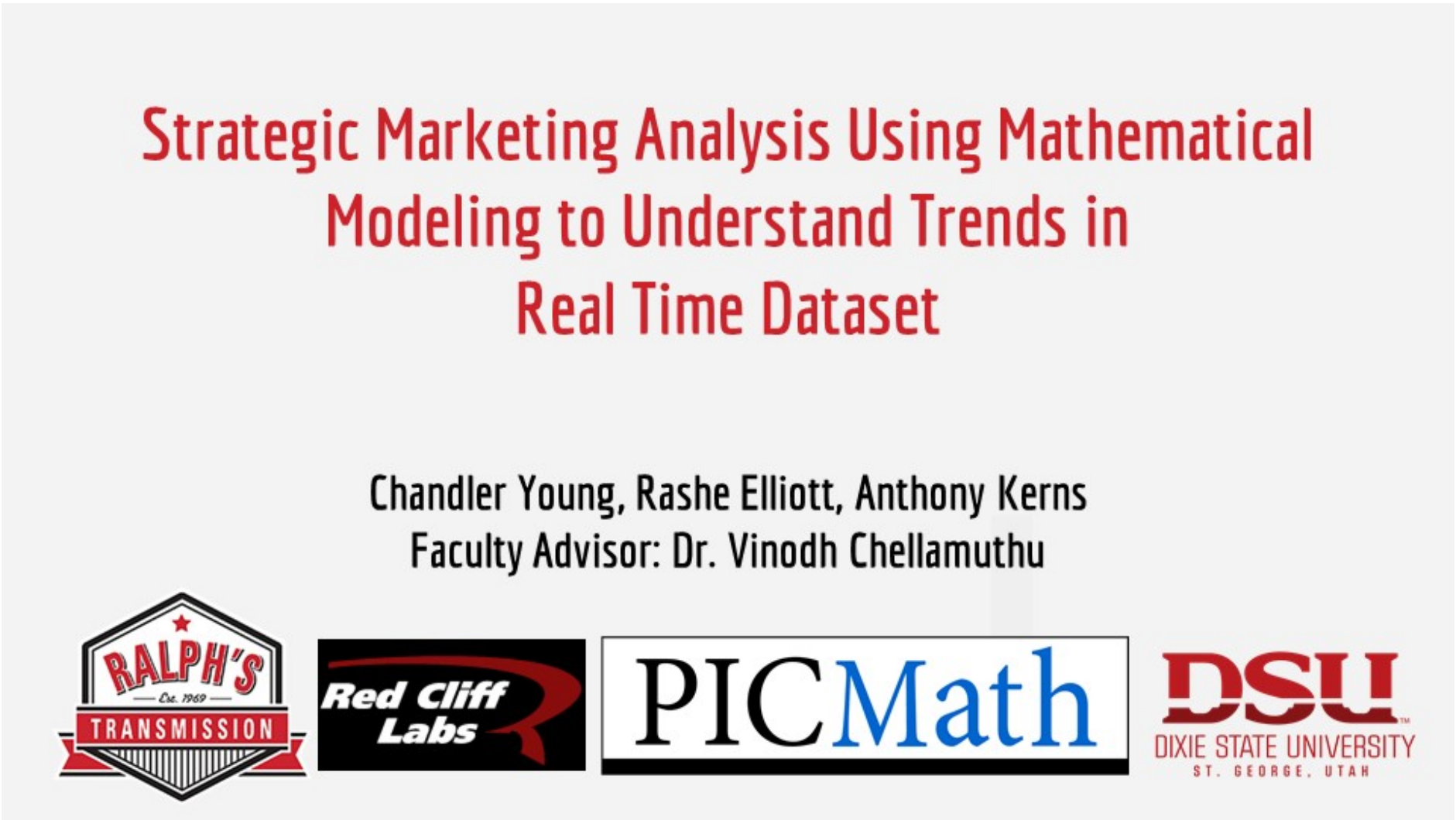
Problems Tackled by our Students - Vinodh

Parks Data



Problems Tackled by our Students - Vinodh

Red Cliff Lab:



Problems Tackled by our Students - Vinodh

Red Cliff Lab:

Problem

Ralph's Transmission posed the following problem to us.

Filter Tool:

- Should show the likelihood of a potential customer to spend money
- This filter should be based on historical data
- This filter should be easy to use.

Marketing Strategy:

- Create a marketing strategy to better allocate resources to increase revenue

Problems Tackled by our Students - Vinodh

Red Cliff Lab:

Algorithm

Categories:

- Average money made for a specific make vs total average money made
- Zip Code/Distance- Top 10 zip codes local to the area
- Make- money made vs total
- Model- money made vs total modified with total model vs Make total

Algorithm:

$$\frac{\$Average_{Make}}{\$Average_{total}} * [Zip\ codes/Distance + \frac{\$_{make}}{total_{make}} + (\frac{\$_{model}}{total_{model}} * \frac{total_{model}}{total_{make}})]$$

Problems Tackled by our Students - Vinodh

Red Cliff Lab:

Scoring and Data Testing Success

Score	Rank
1.6568	HIGH
1.2992	MEDIUM
1.8708	HIGH
1.8449	HIGH
1.3490	MEDIUM
0.8875	LOW
1.3474	MEDIUM
1.9380	HIGH
0.8122	LOW
1.3671	MEDIUM

- **Low Range Score:** 0 - 0.99
- **Medium Range Score:** 1 - 1.59
- **High Range Score:** 1.6+
- Tested the algorithm on various sample sizes, using random sample size as the test data and the remainder as the historical data, while creating and modifying the algorithm

Once confident with results we ran the algorithm on all scrubbed historic data

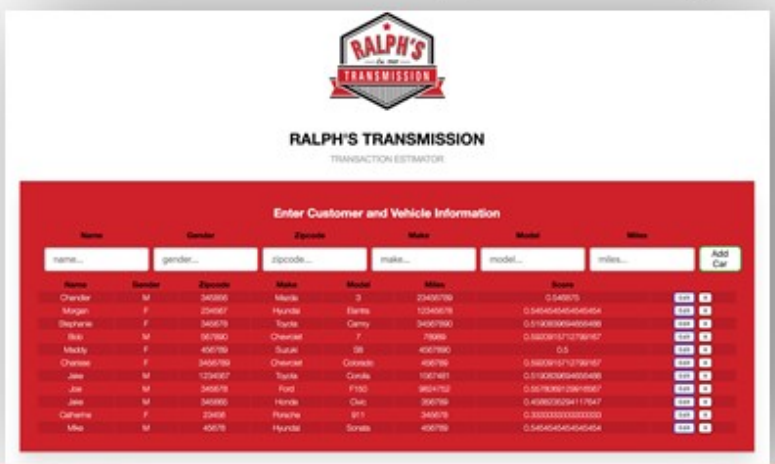
- Approximately **60%** had a successful score to transaction ratio

Problems Tackled by our Students - Vinodh

Red Cliff Lab:

Web Application

- Runs a simple web server and stores all new customer information in an SQL database
- Client Application written in JavaScript, Server Application written in Python (this is where the analyzation of data and math happens, using Pandas package, to generate score)
- See Web Application



Problems Tackled by our Students - Vinodh

Red Cliff Lab:

Marketing Budget Solution

How to Maximize Profit?

- By maximizing the amount of viewers of advertisement

Useful Statistical Data

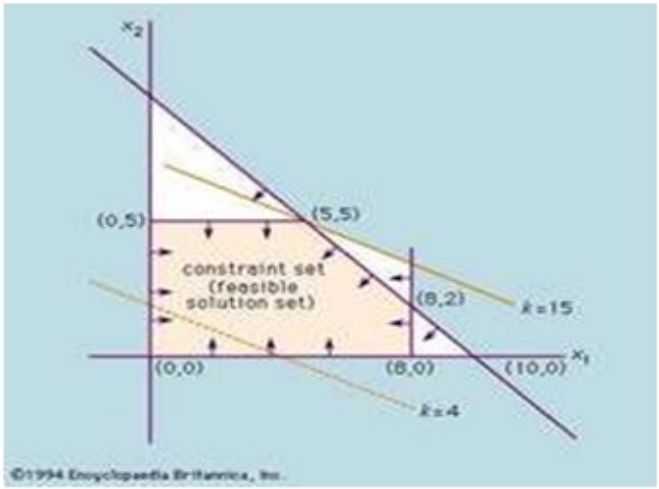
- Top 6 Area Codes: 51.8% of All Potential Customers from Data
- Quantifiable Advertisements: Internet, Radio Commercials, Billboards, Bus Benches, Yellow Pages, Other

Problems Tackled by our Students - Vinodh

Red Cliff Lab:

Blended Simplex Programming Method

- Mathematical modeling technique in which a linear function is maximized or minimized when subjected to various constraints
- Several constraints such as:
  - Geographical Environment
  - Marketing Budget
  - Demographics
  - The Business Marketing Team



Problems Tackled by our Students - Vinodh

Red Cliff Lab:

Marketing Budget Structure

Advertising plan

	Internet	Radio	Bus Benches	Yellow Pages	Billboards	Other	Total
Number ads purchased	15.287	6.369	1.274	1.274	1.274	20.000	25.478
Percentage of budget	**40%**	**25%**	**5%**	**5%**	**5%**	**20%**	100%
Budget by category	$ 8,000	$ 5,000	$ 1,000	$ 1,000	$ 1,000	$ 4,000	$ 20,000

Constraints on numbers of exposures

	Actual exposures		Required exposures	Excess Exposures
Top 6 Zip Codes	121,914	>=	43,160	78,754
	113,223	>=	45,657	67,566
	97,588	>=	8,400	89,188
	86,550	>=	10,250	76,300
	51,408	>=	16,471	34,937
	70,877	>=	57,520	13,357

Objective function to maximize	
Total Excess Exposures	360,102

How We Learned These Tools

Jennifer's Strategies

- MAA Data Science Workshop - intro to Jupyter and Pandas
- Mini-projects using Jupyter/Colab in Finite Math and Statistics
  - Created templates for students to work through
  - Example data from Kaggle
- DataCamp for PIC Math class
  - Free for educators teaching a class
  - Modules to learn Python and R for data analysis
  - Leaderboard with experience points
- Created PIC Math syllabus, calendar, intro survey in LaTeX
- Data analysis for *AMATYC News* in Pandas/Jupyter
- Steal/borrow students' code

Learn by DOING!

Biggest Challenges

Vinodh:

- Mixed levels of mathematical maturity within the class.
- Putting together a team based on their skill set (programming, communication - oral and written skills).
- Mitigating the fear of real-world problem solving and getting them excited and motivated.

Jennifer:

- Lack of knowledge of computing (me!)
- Lack of knowledge of computing (the students)
- Lack of knowledge of data science and analytics (me!)

Both:

- Getting all students on each team actively involved
- TIME CONSTRAINTS

Biggest Rewards

Vinodh:

- Seeing students going from "consumers" to "producers"
- Increased cross-disciplinary collaboration within the university
- Helped to design "hands-on" curriculum within the math department
- Increased opportunities for collaboration with local community
- Small curriculum changes (this class) lead to big innovations:
  - Establishing Modeling and Simulation Hub
  - Creation of Certificate in Modeling and Simulation

Jennifer:

- Interacting with students on deeper level, as collaborators
- Seeing students grow in knowledge and courage
- Personal growth (thinking of me, not the students)
  - "Will this make me braver?"
  - Computing knowledge
  - I now love LaTeX .... I'm never going back!
- Beginning something that will be a long-term plus for my college

Rewards - for Students

Help students grow in:

- Data analysis knowledge - valuable for any major
- Confidence
- Professional communication skills - written and oral
- Problem solving skills
- Critical thinking
- Working and thinking independently

Resulting in:

- More competitive for REUs, scholarships, and internships
- Helps them land first STEM job after graduation

How Can I Use These Ideas?

- Incorporate projects into your normal classes
- Activities for Math/STEM Club
- Look for opportunities at your college to promote:
  - Math Across the Curriculum
  - Computing Across the Curriculum
- Analyze some data as a side project (useful or fun or both)
- Apply for next PIC Math cohort

Timeline if you do PIC Math

- February/March 2021: Apply to PIC Math Program
- Late spring 2021: PIC Math Training Workshop
- Summer and Fall 2021: Find industry partners and projects
- Fall 2021: Recruit students
- Spring 2022: Teach PIC Math class
- Summer 2022: Students present their project at MAA MathFest (or virtual version)

For more information:
`https://www.maa.org/programs-and-communities/professional-development/pic-math`

Some Links Mentioned Earlier

Interactive Plots Created by Students:

- `http://nhmath.lonestar.edu/Faculty/TravisJ/PICMath/ALWest.html`
- `http://nhmath.lonestar.edu/Faculty/TravisJ/PICMath/cameraJuly2.html`

Overleaf Example (LaTeX in the cloud):

- `https://www.overleaf.com/read/vxywktbgdbqt`

Jupyter/Colab (for plots and analysis)

- `https://colab.research.google.com`

Thank you! Any questions?

- Jennifer Travis
  Jennifer.Travis@LoneStar.edu
- Vinodh Chellamuthu
  Vinodh.Chellamuthu@dixie.edu