Syracuse University
Justin Pate

PORTFOLIO OF LEARNING OBJECTIVES

THE iSCHOOL
Syracuse University

APPLIED DATA SCIENCE

DECEMBER 26, 2019

Introduction—Justin Pate

I received my bachelors while overseas in the Air Force from American Military University. After returning home, I entered into an NC programming role which eventually lead to precision measurement and automation integration.

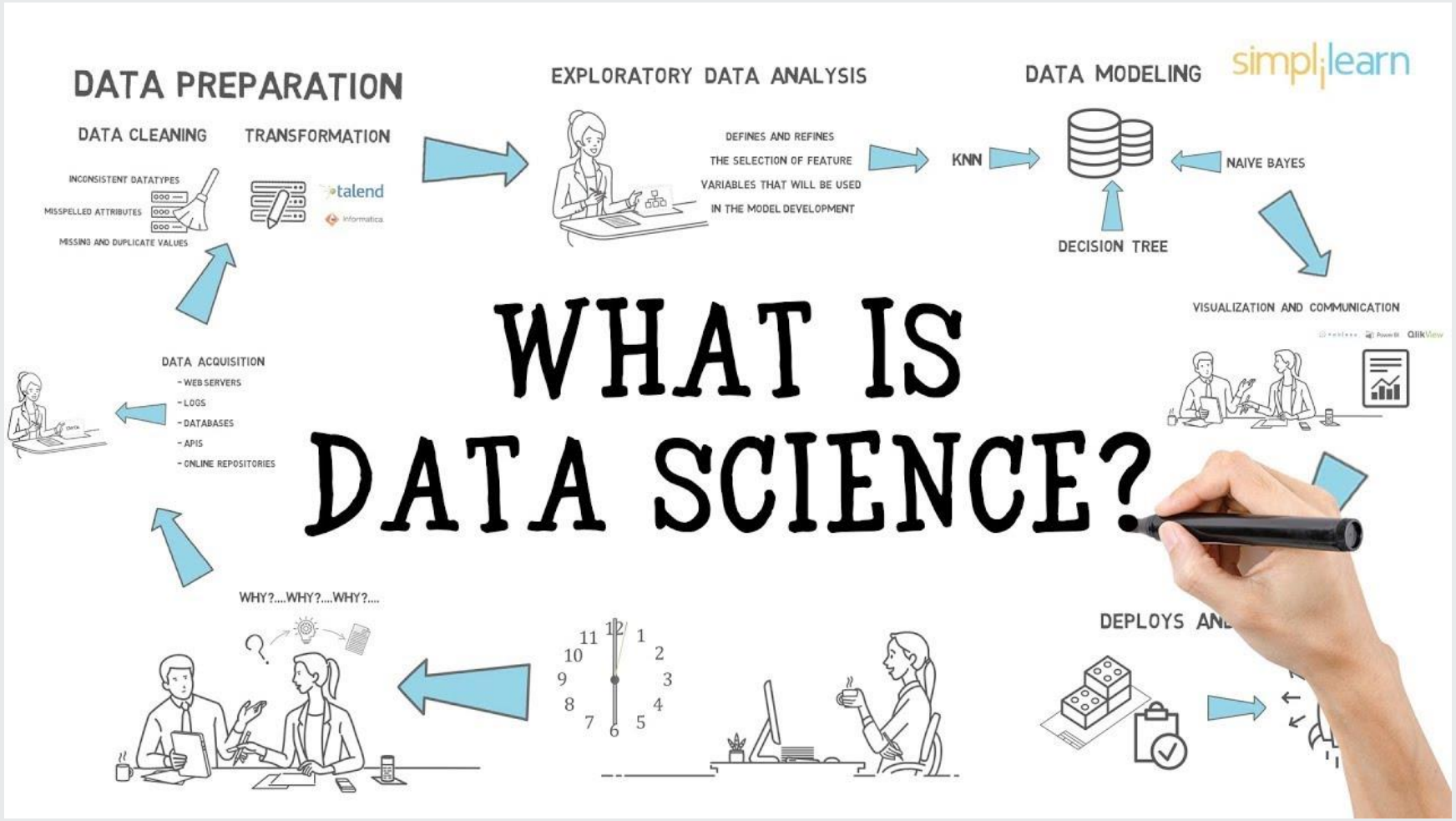
This has been an extremely rewarding career. It is the perfect mix of hands on trouble shooting with data and integrating custom technical tools the improve a fast paced manufacturing environment.

U.S. AIR FORCE

AMU

THE iSCHOOL
Syracuse University

What is Data science?

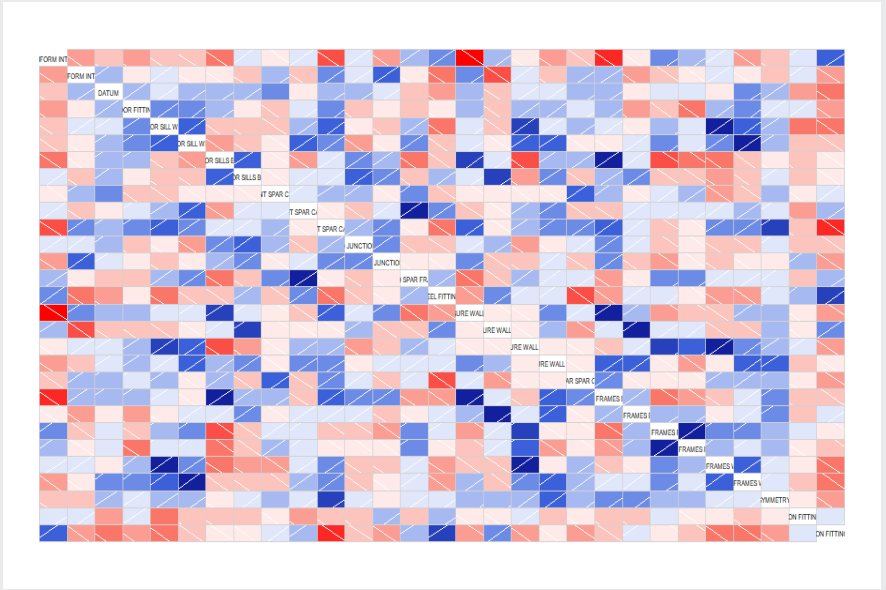


In my own words, data science is harnessing a data intensive world to gain a better understanding of human behavior as well as mechanical prediction.

Project Based learning-Wing Measure data

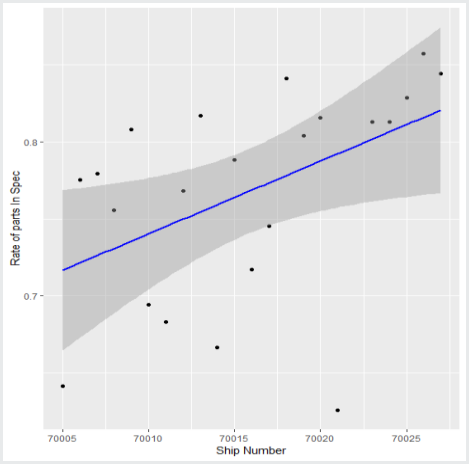
Currently, aerospace structures supplier “We Make Wings” supplies the wings for a jet. “Final Product Builder” is the customer. Final Product builder has given some very aggressive tolerances for the scope and large size of the structure. Base on approximately 20 units, enough data has been collected on the build process to evaluate. The company has now enlisted a team of Graduate students from Syracuse to identify opportunities in the data.

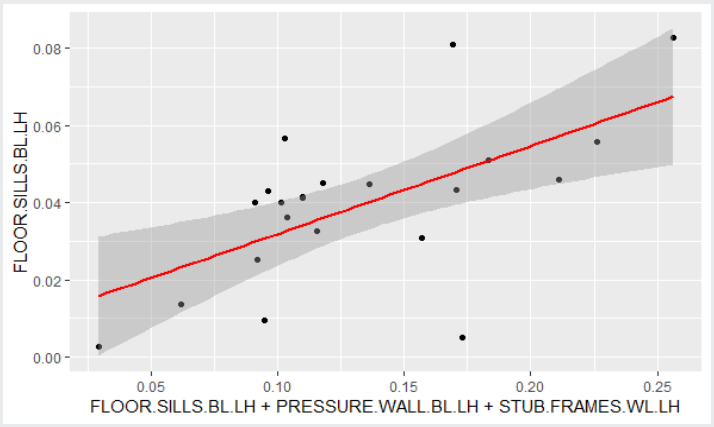
1. Which of the features are related via statistical correlation?
2. Can a valid model be created based on the data?
3. Is the build process improving or degrading over time?
4. What processes are we good at?
5. What processes are we bad at?
6. Are there any obvious outliers in the data?



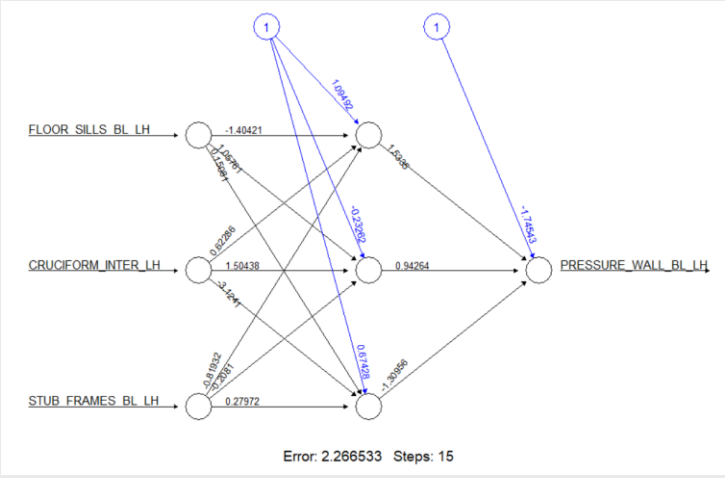
Dark red = strong negative correlation

Dark blue = strong positive correlation





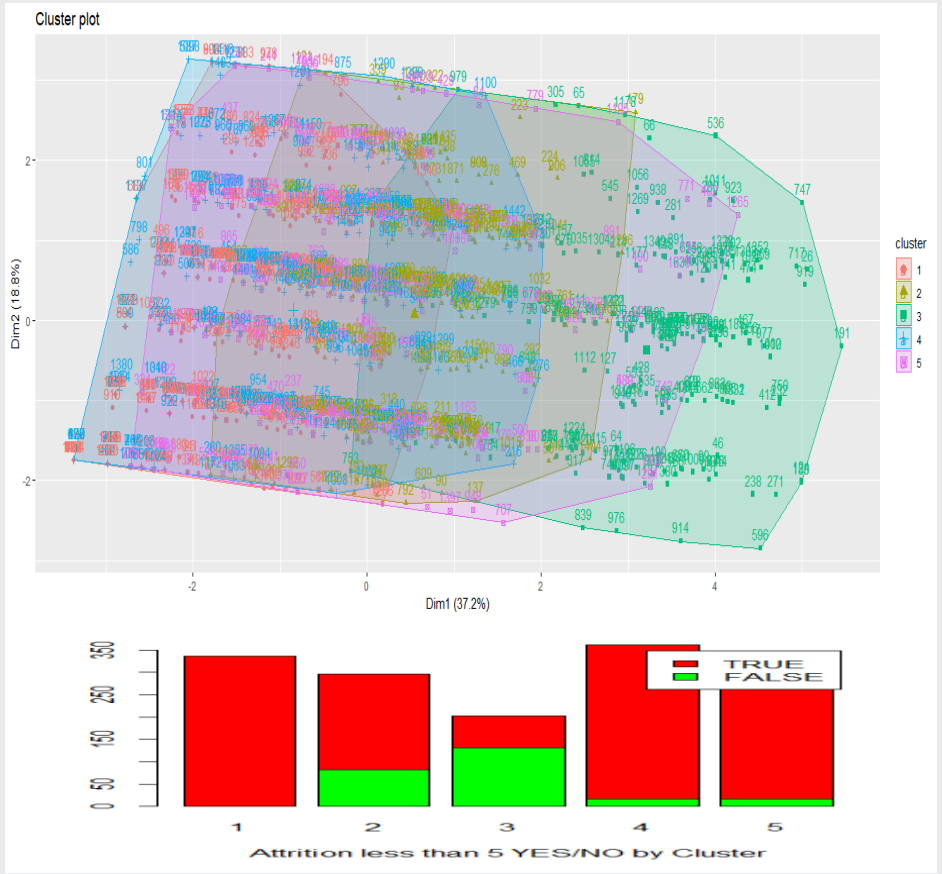
	[,1]
error	2.26653348
reached.threshold	0.09073234
steps	15.00000000
Intercept.to.1layhid1	1.09491709
FLOOR_SILLS_BL_LH.to.1layhid1	-1.40421027
CRUCIFORM_INTER_LH.to.1layhid1	0.62286339
STUB_FRAMES_BL_LH.to.1layhid1	-0.81932087
Intercept.to.1layhid2	-0.23262096
FLOOR_SILLS_BL_LH.to.1layhid2	1.05761043
CRUCIFORM_INTER_LH.to.1layhid2	1.50437538
STUB_FRAMES_BL_LH.to.1layhid2	-0.20810186
Intercept.to.1layhid3	0.67427960
FLOOR_SILLS_BL_LH.to.1layhid3	0.15081302
CRUCIFORM_INTER_LH.to.1layhid3	-3.12410068
STUB_FRAMES_BL_LH.to.1layhid3	0.27972210
Intercept.to.PRESSURE_WALL_BL_LH	-1.74542678
1layhid1.to.PRESSURE_WALL_BL_LH	1.53349819
1layhid2.to.PRESSURE_WALL_BL_LH	0.94263933
1layhid3.to.PRESSURE_WALL_BL_LH	-1.30956327

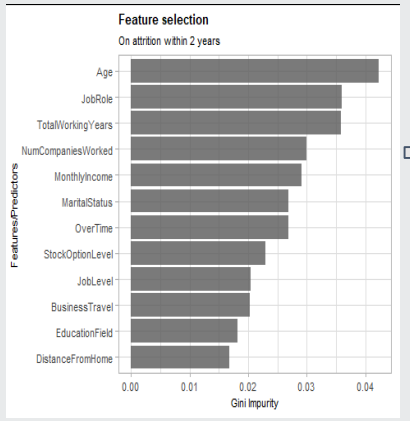


Conclusoin resulted in a neural NET model that predicted within 75% accuracy if Keel beam down stream features would be in tolerance.

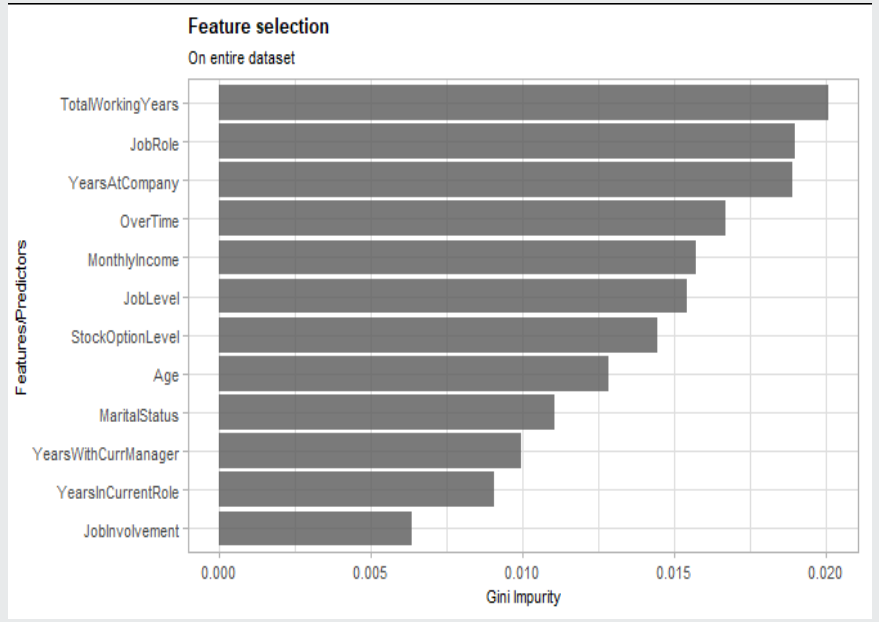
Project Based learning-Employee Attrition Analysis

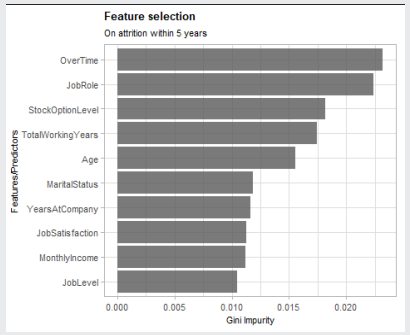
Employees are both an investment and an appreciating asset. They are an investment because of all the time and effort it takes to find, train and develop them. And they appreciate in value because over time they become more efficient at their work, they can identify valuable relationships between business processes and identify ways to eliminate waste.



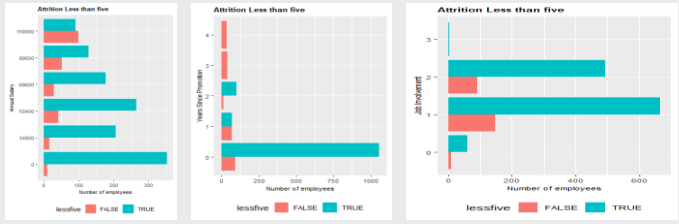


78%





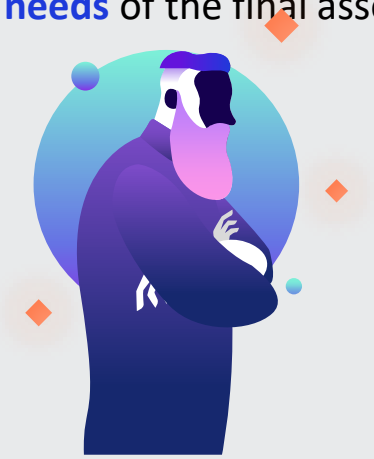
83%



Project Based learning-Manufacturing Manpower

PROBLEM STATEMENT

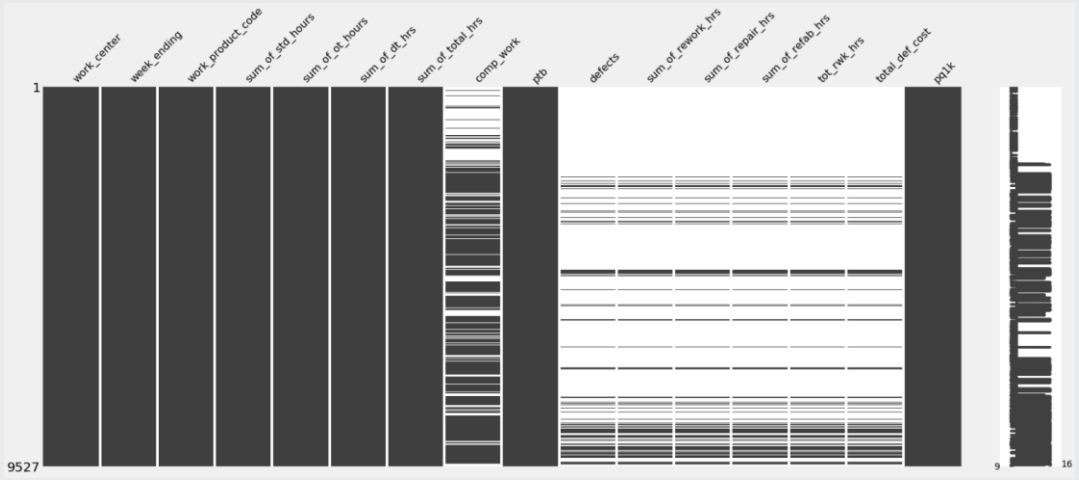
In **large manufacturing environments**, the need to **predict** performance trends and quality rework is imperative to maintain a schedule that **meets the needs** of the final assembly ine of the customer. This need is abundantly present in **aerospace**.



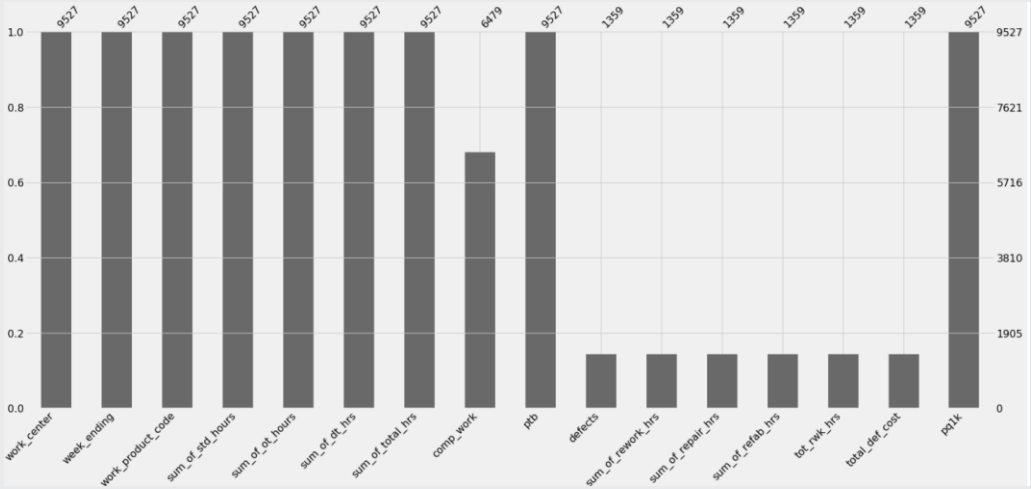
	WORKCENTER	WEEK_ENDING	WKOR_PRODUCT_CODE	SumOfstdhours	SumOfOThours	SumOfdthrs	umOftotalhrs	comp_work	PTB	defects	SumOfREWORK_HRS	SumOfREPAIR_HRS	SumOfREFAB_HRS	totrwkhrs	total_def_cost
8560	J0921201	3/9/2019	WA	65.9	17.0	0.0	82.9	82.870	99.96%	NaN	NaN	NaN	NaN	NaN	NaN
2185	J0222101	10/6/2018	BW	37.1	32.5	0.0	69.6	69.572	99.96%	21.0	0.0	4.0	0.0	4.0	686.0
5002	J0747505	10/20/2018	BW	143.2	32.3	8.0	183.5	183.404	99.95%	1.0	1.0	0.0	0.0	1.0	126.0
5562	J0747703	8/25/2018	BW	26.7	0.0	0.0	26.7	26.675	99.91%	NaN	NaN	NaN	NaN	NaN	NaN
4684	J0747221	11/10/2018	BW	82.8	21.8	26.1	130.7	130.534	99.87%	NaN	NaN	NaN	NaN	NaN	NaN

A performance to budget calculation and defects normalized by completed work were added to the dataset.

FULL DATASET COMPLETENESS

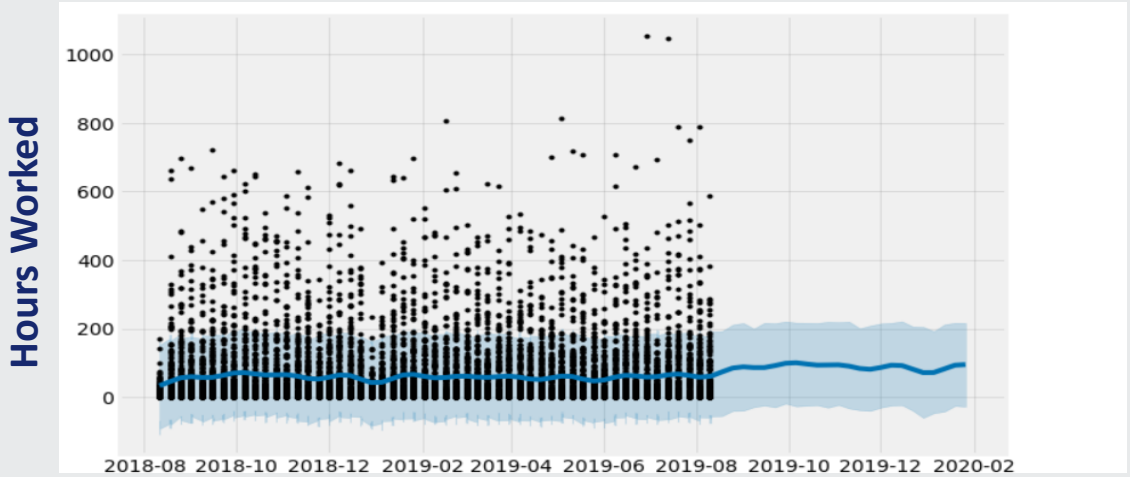


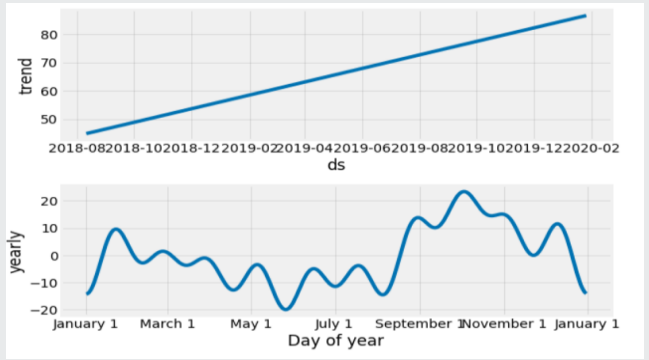
COMPLETENESS BY COUNTS



Project Based learning-Manufacturing Manpower (cont)

ANALYSIS HOURS WORKED





Trend over **years**

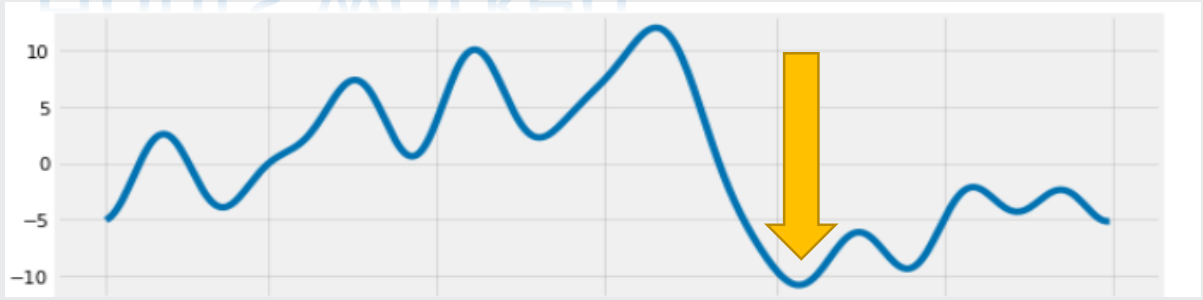
Trend over **months**

- Over all **increase** of hours worked in both the actual data and forecast
- **Increase** in hours worked in the last three months of the year
- **Decrease** in hours worked immediately prior to New Year's Day
- Actual hours worked **decrease** in the summer between May and July

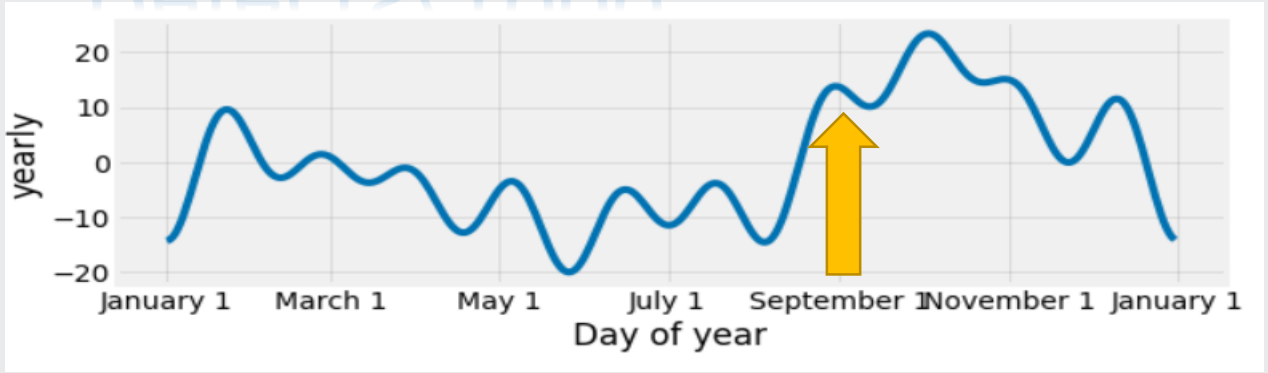
Project Based learning-Manufacturing Manpower (cont)

RECOMMENDATION

Hours worked

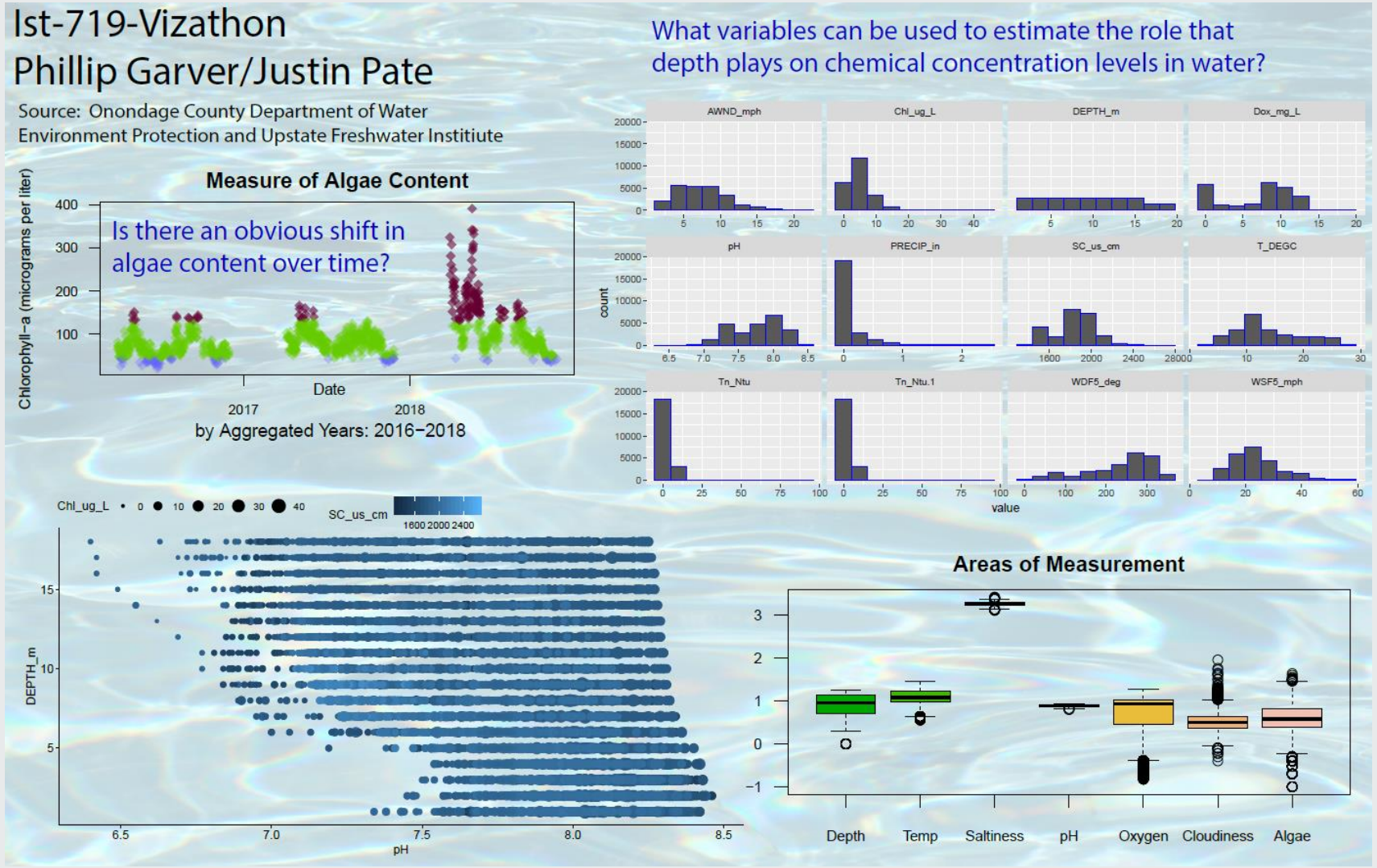


Defects/1000



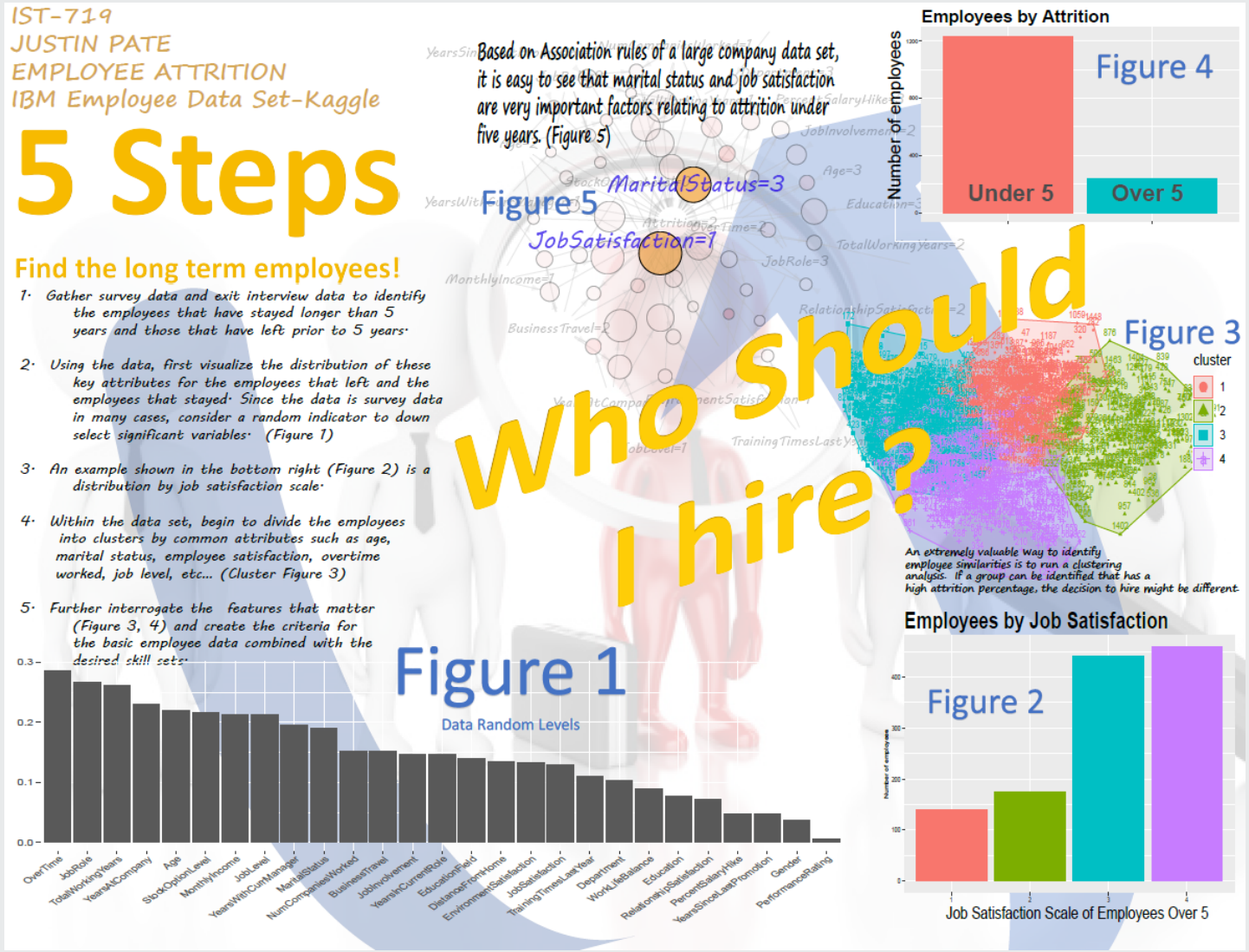
1. **Increase man-power**
   - To counter the effect defects will have and to supplement the estimated decline.
2. **Manage vacation schedules**
   - Request some employees to schedule vacations earlier so that they can be coordinated to account for the anticipated drop in production quality
3. **Apply above steps to other forecasted periods where working hours are low and defects are high**
   - Or where negative trends are predicted to negatively impact production
4. **Survey**
   - Survey the employees to try to understand relationship of employee schedules to seasonal variation

Project Based learning-Vizathon Lake Water Quality



- This is an assignment that challenged the data scientist to perform under pressure.
- The task was to take a dataframe in the time allotted in one class, coordinate with another classmate. Produce R code to create the visualizations, create an adobe illustrator file with layers, and produce a vector imaged PDF, zoomable without loss in clarity.
- The result is something to be proud of.

Project Based learning-FINAL POSTER—EMPLOYEE DATA

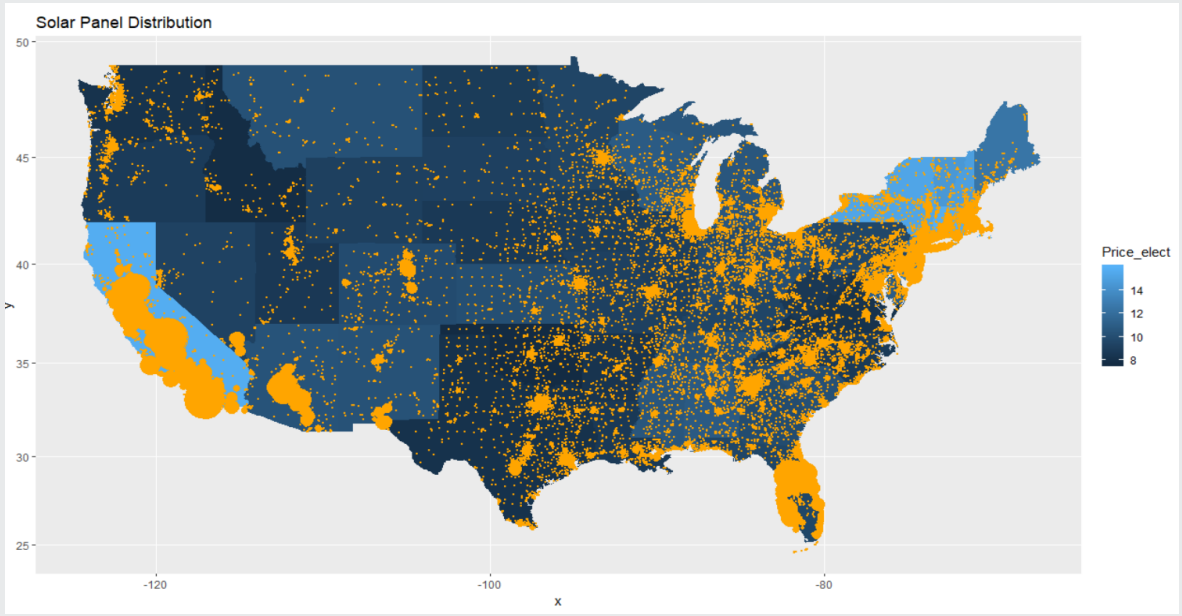


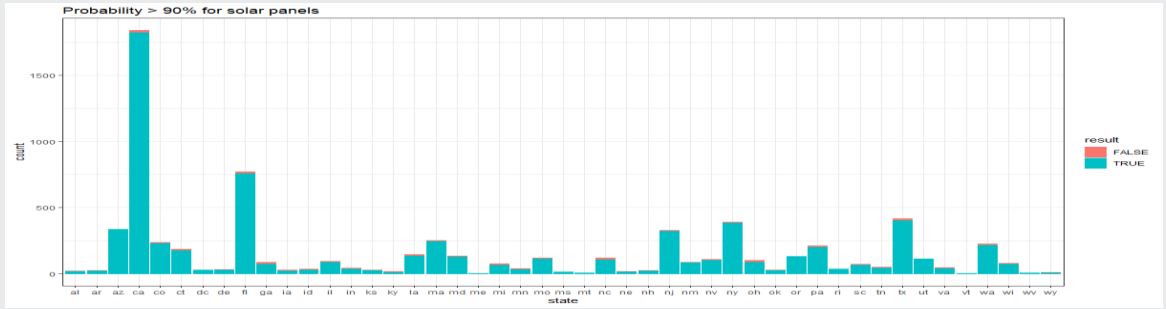
- This project took another look at the employee attrition dataset from Kaggle.
- In this effort, the goal was to tell a story to Human resource professionals regarding the steps to hire people who will stay longer than five years.
- The result is a fantastic looking poster with instructional steps.

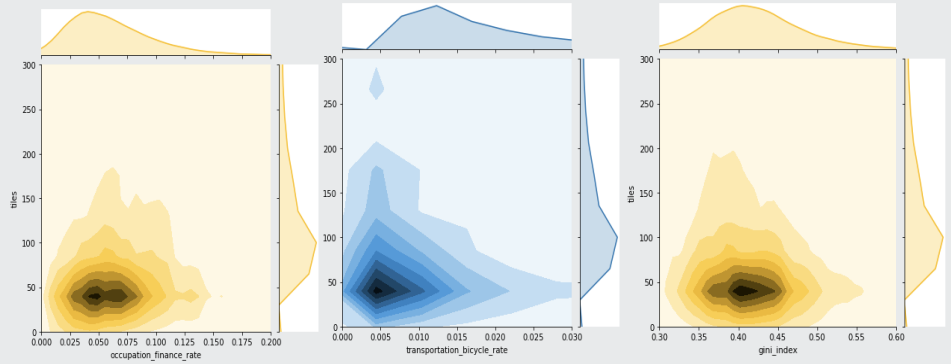
Project Based learning-Solar Panel Sales

We've seen solar installation data over electricity pricing.

The dynamics at play are not clear. An analysis is needed to point out areas where conditions are optimal.





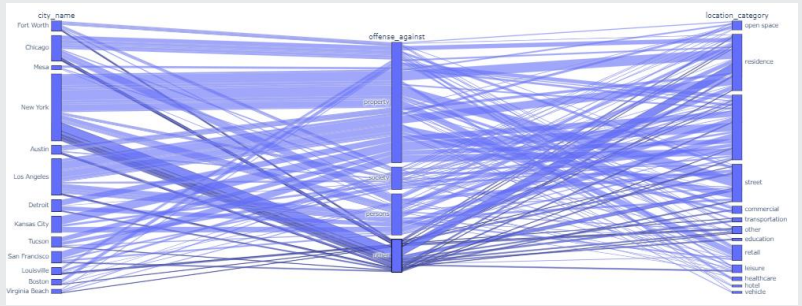


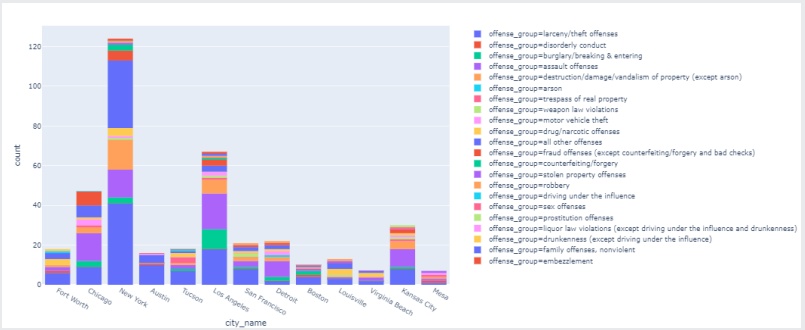
Conclusion is a dataset that can be used by solar panel companies as a marketing campaign in regions most likely to have a solar panel market.

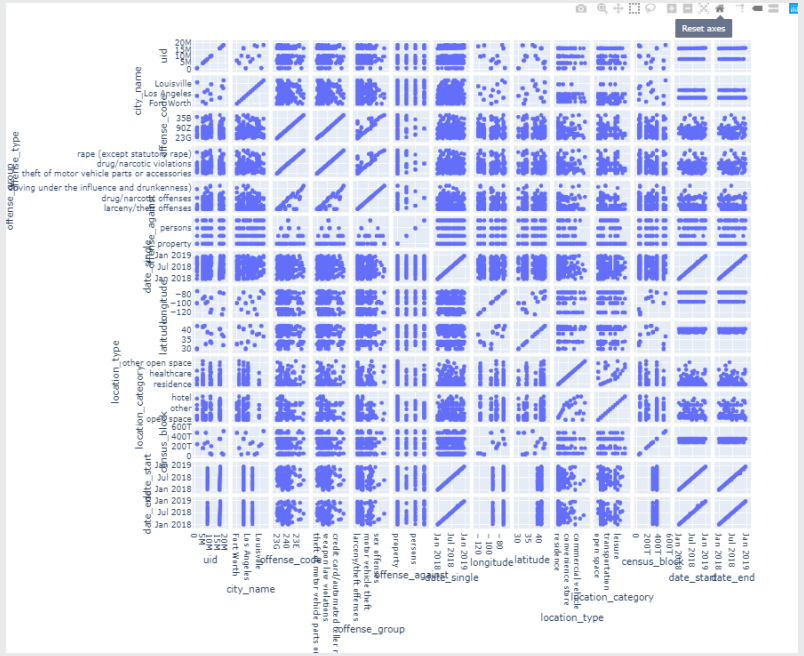
Project Based learning-Major City Crime data

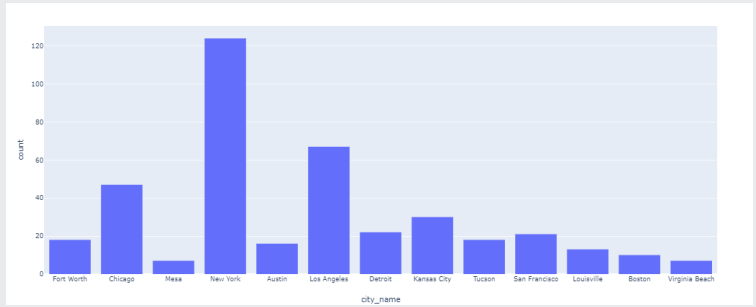
Using variations of Visual Data representation and aggregates, conclusions can be made regarding city and crime type and density. The first primary tool used is a mapping tool where a function is created in the program that controls the initial map zoom and the re-centering of the map based on the data to be displayed. The second primary tool used is plotly.express which is a phenomenal interactive visualization tool that can be imported into python.

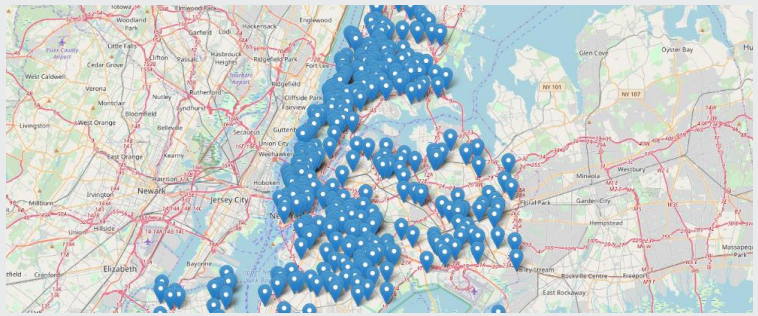
Interactive demonstration of the below visualizations to be given NOW!. All visualizations can also be found in write up.









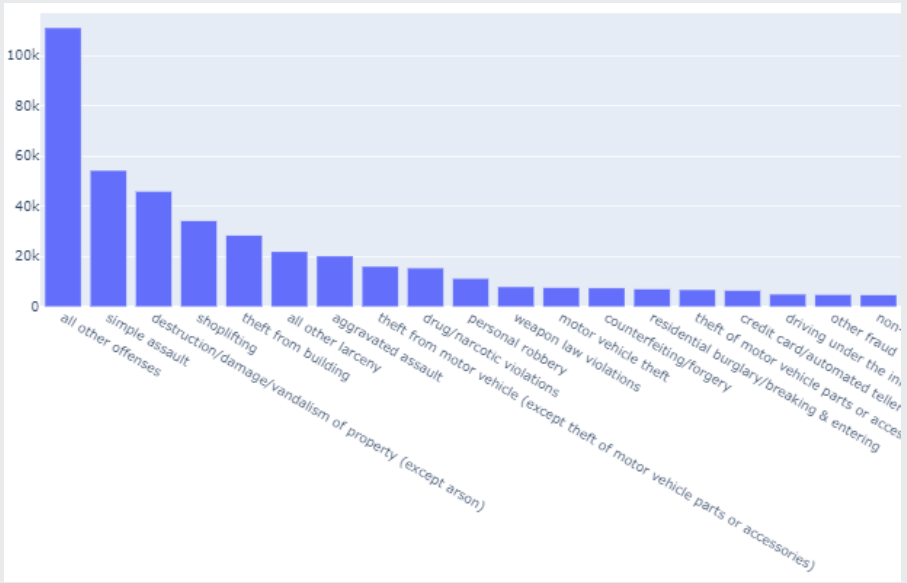


Project Based learning-Major City Crime data (cont)

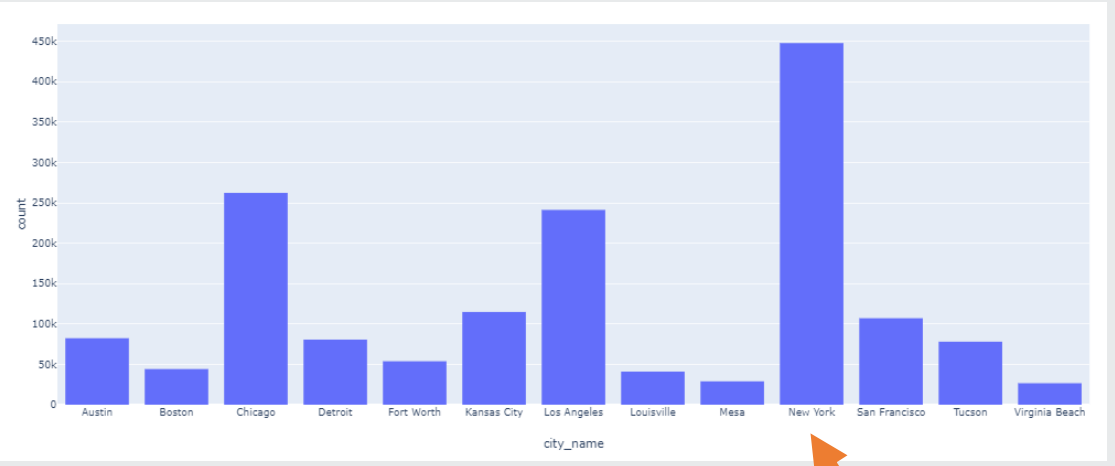
Some simple aggregation and visual analysis below allow some very confident and quick conclusions.

New York clearly has the highest non-normalized crime rate in the dataset that is analyzed.

A second break down could be by crime type within the city of new York.



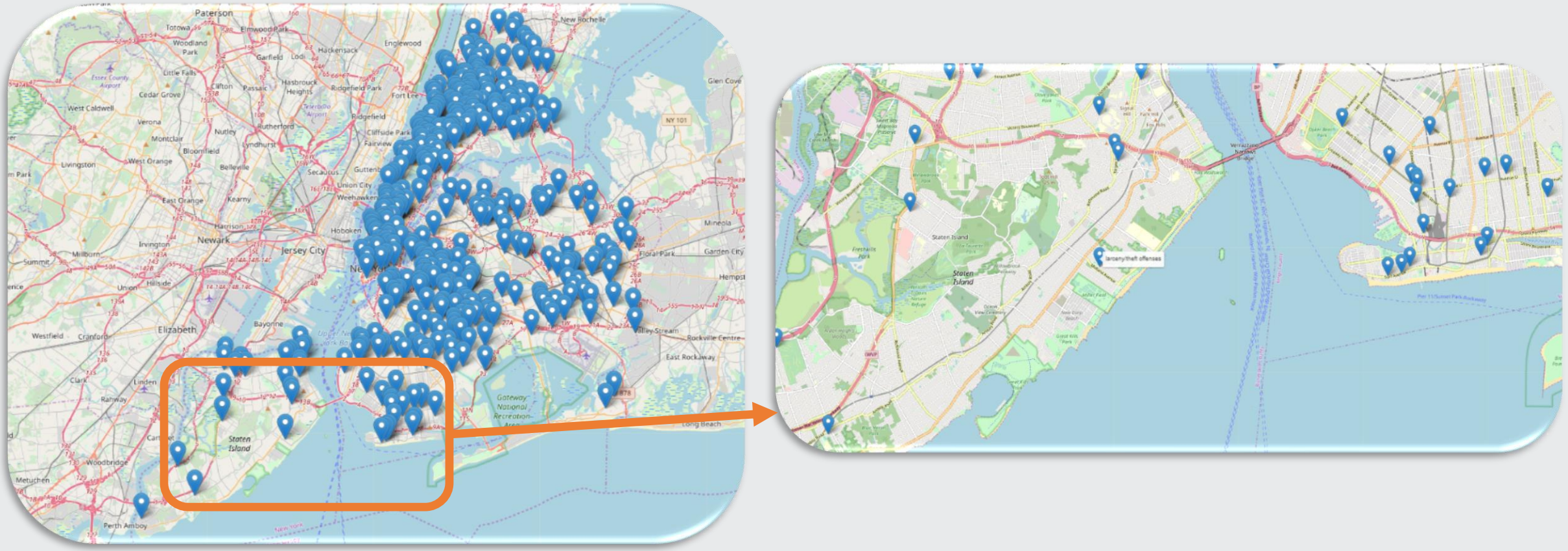
city_name	
Austin	82353
Boston	44165
Chicago	262258
Detroit	80618
Fort Worth	53819
Kansas City	114889
Los Angeles	241220
Louisville	41008
Mesa	28947
New York	447766
San Francisco	107095
Tucson	78033
Virginia Beach	26618



Project Based learning-Major City Crime data (cont)

But what if I still wanted to live in New York?

Based on the tool below, a location without a large crime density could be identified.



- Thank you Syracuse

- What is next?