



# HR DATA ANALYTICS PROJECT



City of Tacoma  
WASHINGTON



Jamilya Shokanova



Jakob Jarecki



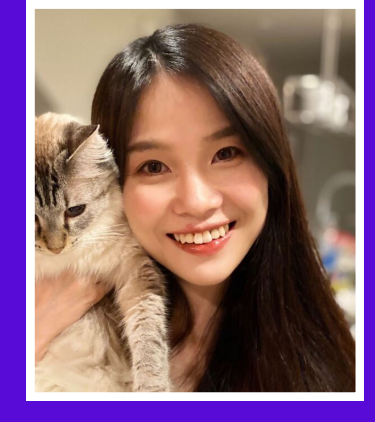
Caitie Johnson



Sean Connolly



Mrudula Achanta

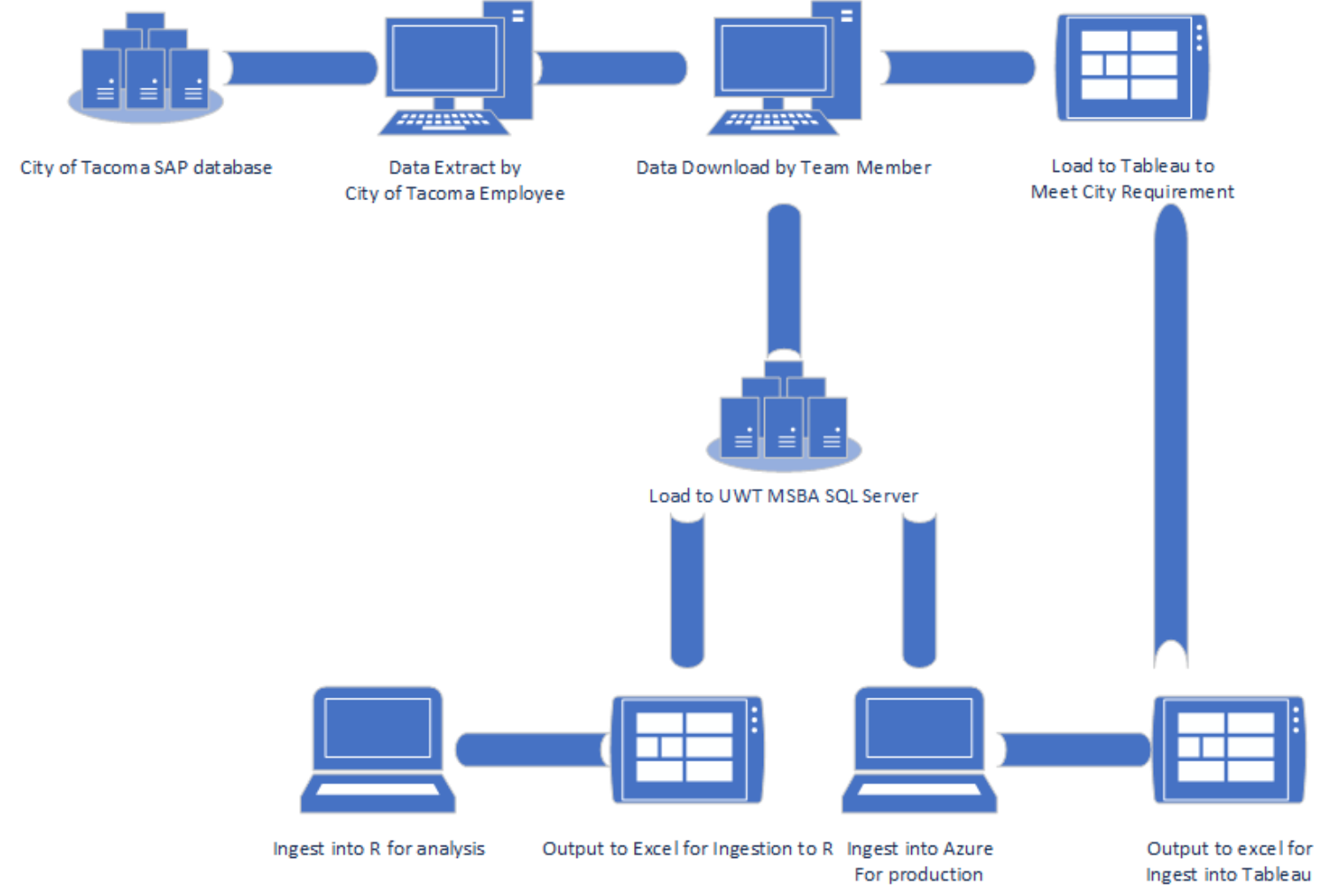


Can Peng

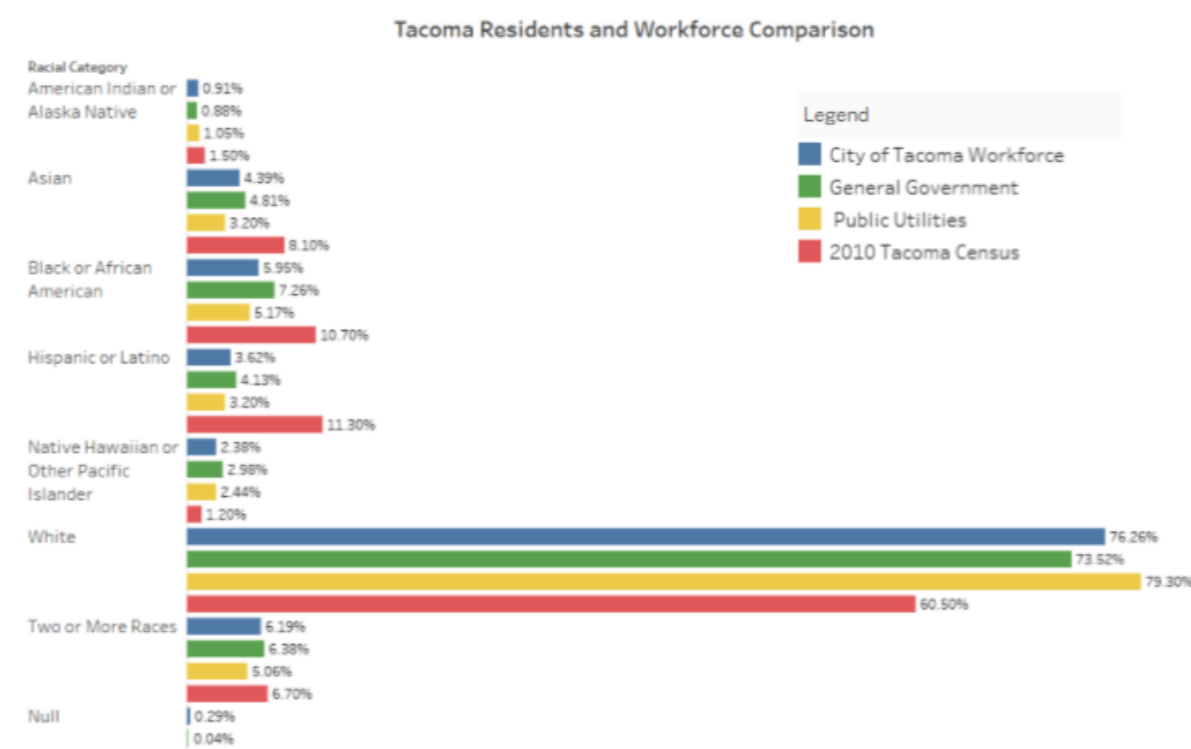
## Project Background and Goals

The HR department is looking to partner with our consultancy group to help better understand their large unstructured data, to provide insights about recruitment and hiring, equity and diversity and workforce planning.

**Goal:** To create a useable, well structured, and predictable visualization tool. This will allow the department to automate reporting and provide insights that will contribute to their ongoing efforts of attracting, developing and retaining a talented workforce .



## Analytics Techniques and Technology



Descriptive and Diagnostic analysis were completed in Excel and Tableau. A proof of concept – Tableau dashboard was created with information pertaining to training, salary and demographic data. In addition, graphs that were being updated manually by the department for EEO reporting were also incorporated into the dashboard which will save time and lead to less data entry errors.



A predictive attrition model was developed in R studio then later transferred into Azure machine learning studio. Several models were tested including Logistic regression, random forest and boosted decision tree.



```

call: glm()
deviance Residuals:
    Min       1Q   Median       3Q      Max
-1.0020  -0.3841  -0.2504  -0.2178   3.1064

Coefficients:
(Intercept)            Estimate Std. Error z value Pr(<|z|)
-0.207756003      0.470906676  -0.441    0.65908
-0.040989525      0.008432521  -4.861    0.0000117 ***
-0.000014913      0.0000004549  -3.212    0.00132 **
 1.073022890      0.736454381    1.457    0.14511
-0.541190665      0.588480465   -0.920    0.35776
 0.786624293      0.314677761    2.500    0.01243 *
-0.266284003      0.400068848   -0.666    0.50569
-0.974226102      0.391144172   -2.491    0.01275 *
 0.029742138      0.347350057    0.086    0.93176
 0.022064027      0.365828361    0.060    0.95188
 0.449425094      0.244833681    1.836    0.06641 .

---
signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 1136.9  on 2516 degrees of freedom
Residual deviance: 1052.2  on 2506 degrees of freedom
AIC: 1074.2

Number of Fisher Scoring Iterations: 6
  
```

## Recommendations & Conclusions

Make data integrity a priority – standardize data types and formats to make joining data sets easier.

Update data on a regular cadence and make the transition to SQL server instead of data entry into Excel.



Start tracking additional metrics that could be used in the predictive model, such as work center data- office locations of each employee and distance traveled for work.

Using employee engagement survey data, we were able to utilize the Azure Text analytics API to conduct key phrase extraction, sentiment analysis and text categorization. We were able to identify actionable insights and summarize the data into a more digestible format.

