Uncovering The Great Resignation

Priyanka Ghosh Dastidar

Kunal Sanghvi

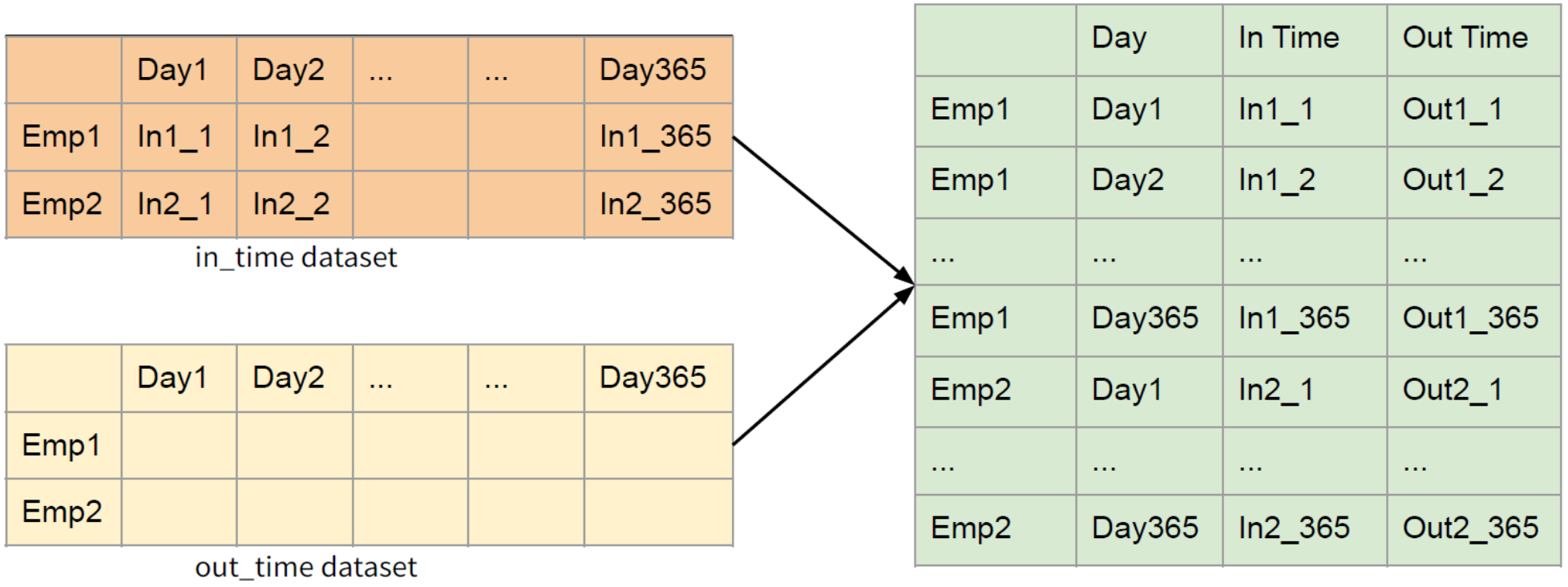
Nithyashree Senguttuvan

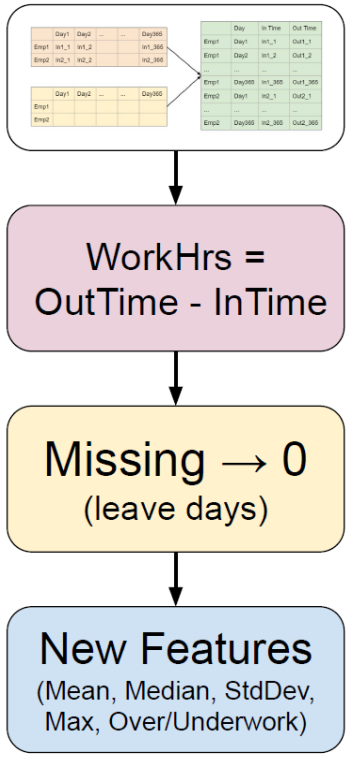
Datasets

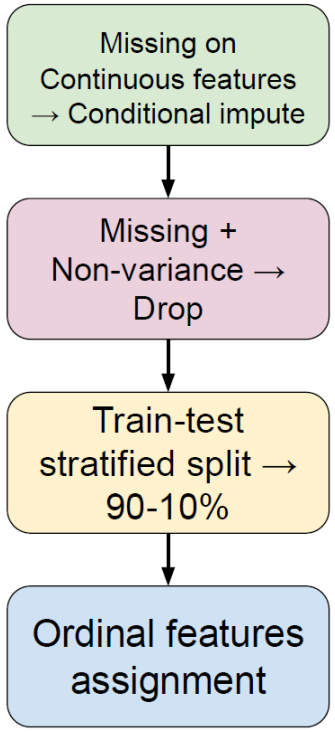
- Source: https://www.bls.gov/

File name	Description
data_dictionary	Definition of all variables available for study
employee_survey_data	Employee survey inputs
manager_survey_data	Manager survey inputs
general_data	Employee descriptive variables and attrition
in_time	Employee clock-in times
out_time	Employee clock-out times

Exploratory Data Analysis (EDA)







Features with missing data

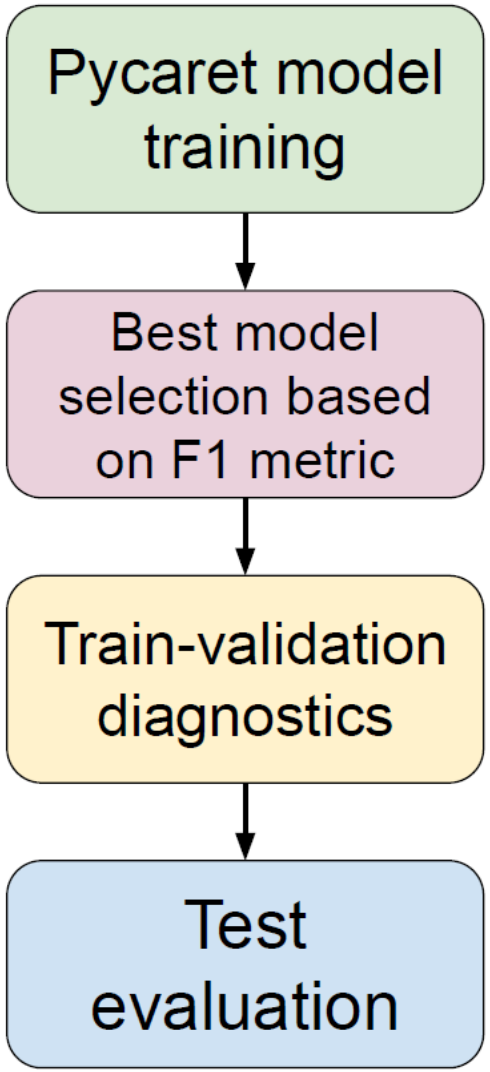
col	num_missing	pct_missing
EnvironmentSatisfaction	25	0.567
JobSatisfaction	20	0.454
WorkLifeBalance	38	0.862
NumCompaniesWorked	19	0.431
TotalWorkingYears	9	0.204

Non-variance features

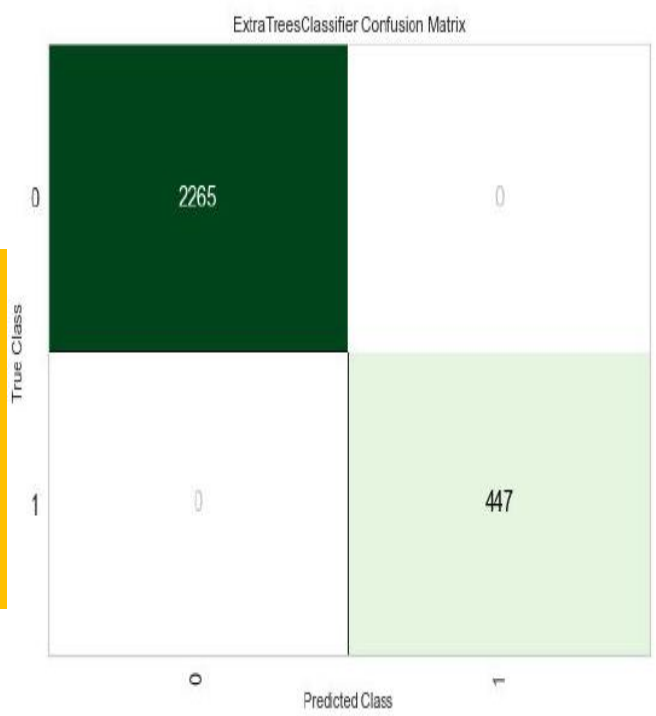
```
non_variance_cols:
EmployeeCount     1
Over18            1
StandardHours     1
```

Feature Engineering

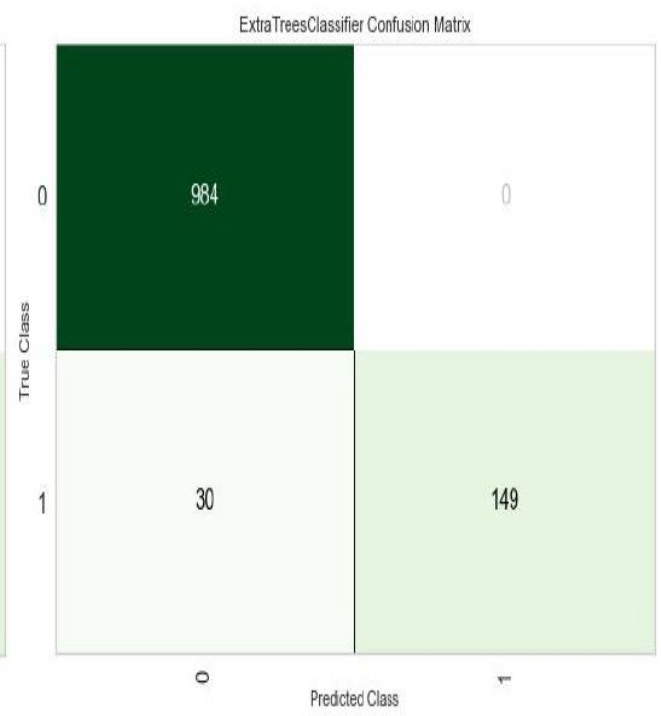
Baseline Model



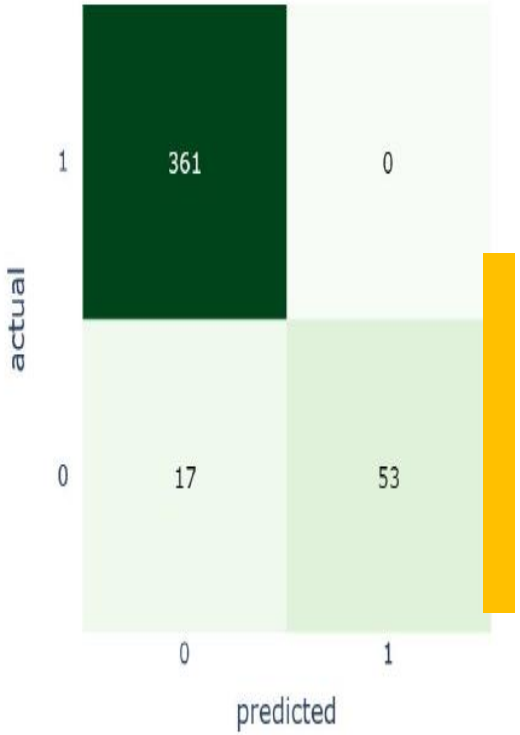
	Model	Accuracy	AUC	Recall	Prec.	F1	Kappa	MCC
et	Extra Trees Classifier	0.9757	0.9899	0.8615	0.9898	0.9200	0.9058	0.9096
gpc	Gaussian Process Classifier	0.9631	0.9313	0.8547	0.9189	0.8837	0.8619	0.8640
rbfsvm	SVM - Radial Kernel	0.9001	0.9835	0.3938	1.0000	0.5633	0.5191	0.5921
rf	Random Forest Classifier	0.8879	0.9451	0.3467	0.9510	0.5031	0.4552	0.5315
qda	Quadratic Discriminant Analysis	0.8293	0.7926	0.4697	0.4851	0.4753	0.3738	0.3749
nb	Naive Bayes	0.8326	0.7263	0.2686	0.4873	0.3451	0.2585	0.2741
lda	Linear Discriminant Analysis	0.8473	0.7966	0.2346	0.6001	0.3353	0.2674	0.3062
knn	K Neighbors Classifier	0.8197	0.7992	0.2353	0.4213	0.2969	0.2044	0.2175
ridge	Ridge Classifier	0.8451	0.0000	0.1025	0.6746	0.1770	0.1419	0.2208
lr	Logistic Regression	0.8381	0.7122	0.0649	0.3552	0.1096	0.0828	0.1211
mlp	MLP Classifier	0.7033	0.5330	0.2022	0.0835	0.0616	0.0036	0.0138
dt	Decision Tree Classifier	0.8293	0.5000	0.0089	0.2091	0.0148	0.0019	0.0215
ada	Ada Boost Classifier	0.8241	0.5720	0.0089	0.0105	0.0096	-0.0066	-0.0066
svm	SVM - Linear Kernel	0.8352	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
gbc	Gradient Boosting Classifier	0.8352	0.6476	0.0000	0.0000	0.0000	0.0000	0.0000
lightgbm	Light Gradient Boosting Machine	0.8352	0.5128	0.0000	0.0000	0.0000	0.0000	0.0000



training



validation

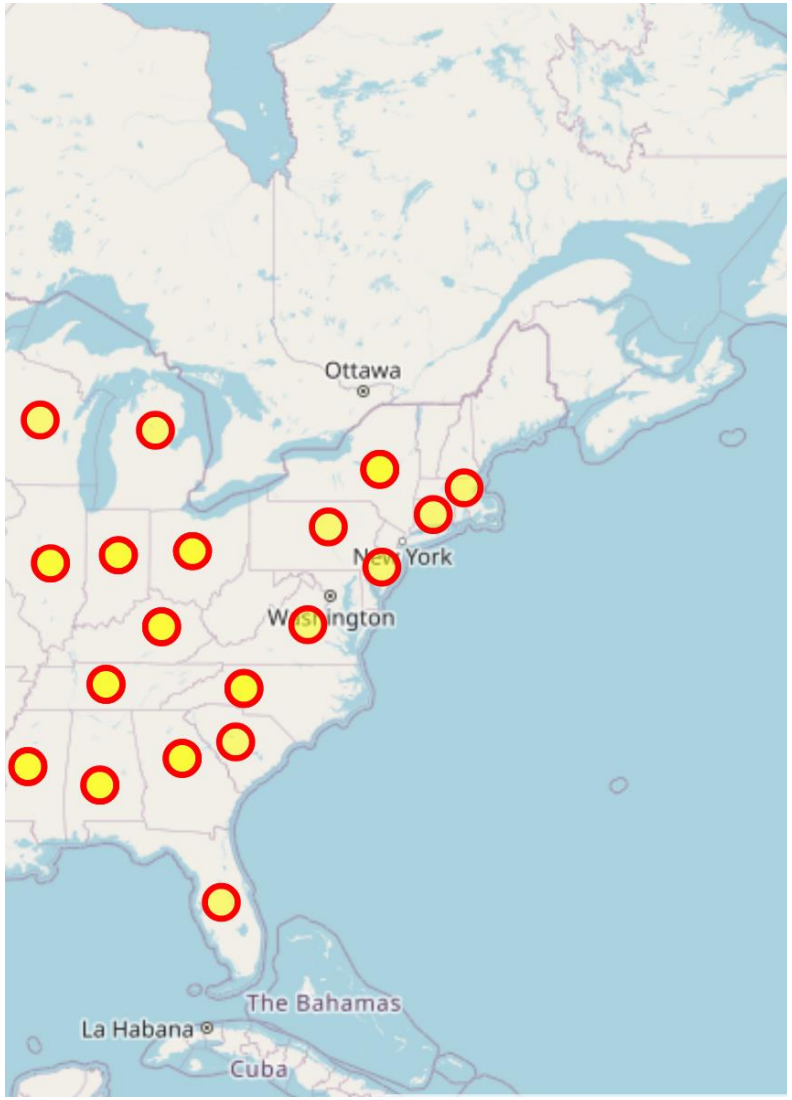


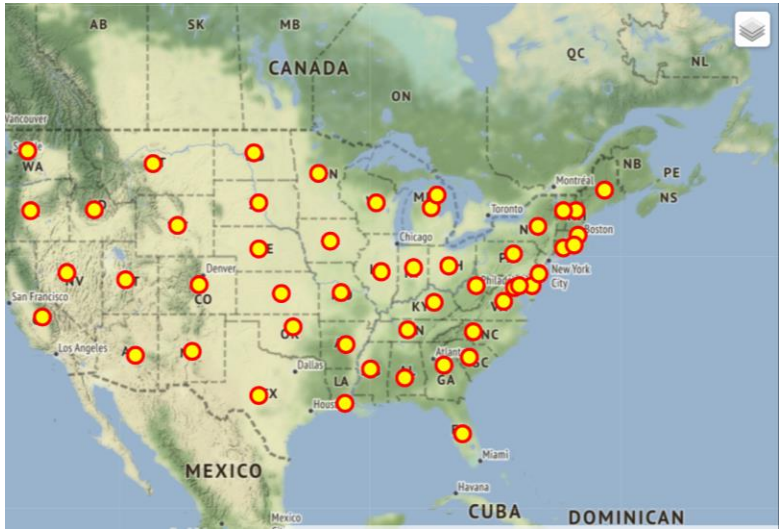
testing

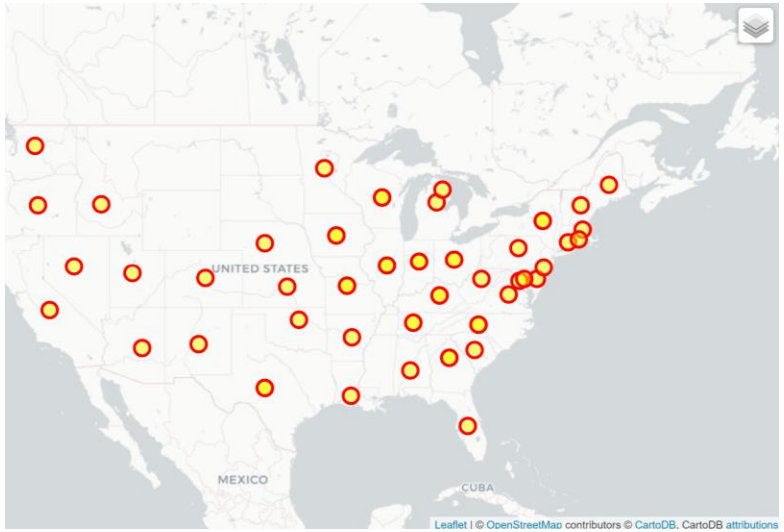
Confusion Matrix

VISUALIZATION INSIGHTS - PYTHON CODE

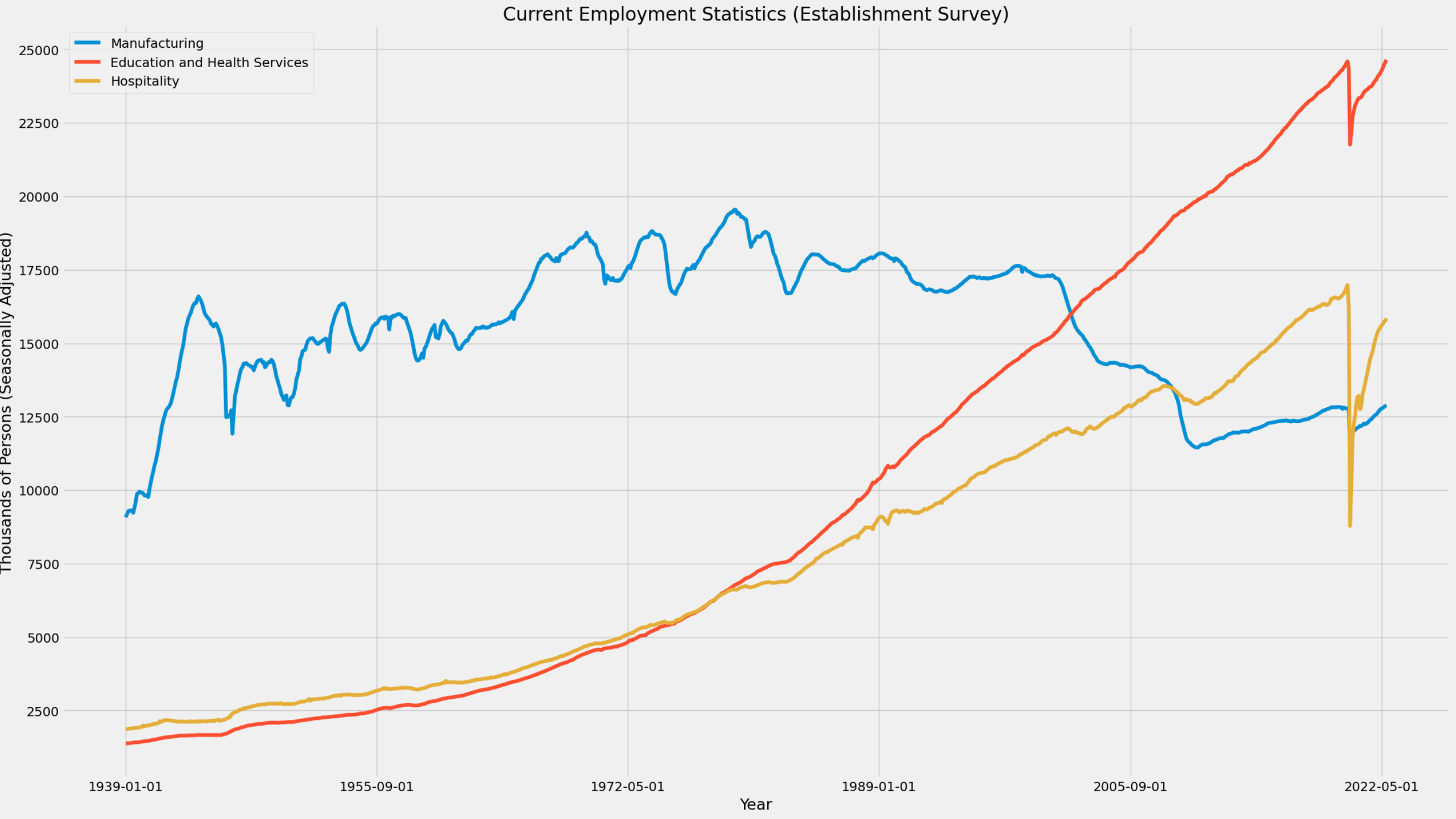
- Geo-spatial visualization of attrition - US states using the Labor Data
- Comparison of Employee Survey on Education Field, Job Involvement, Job Satisfaction, Environment Satisfaction, Relationship Satisfaction and Work Life Balance.
- Industries Most Affected
- Industries Least Affected

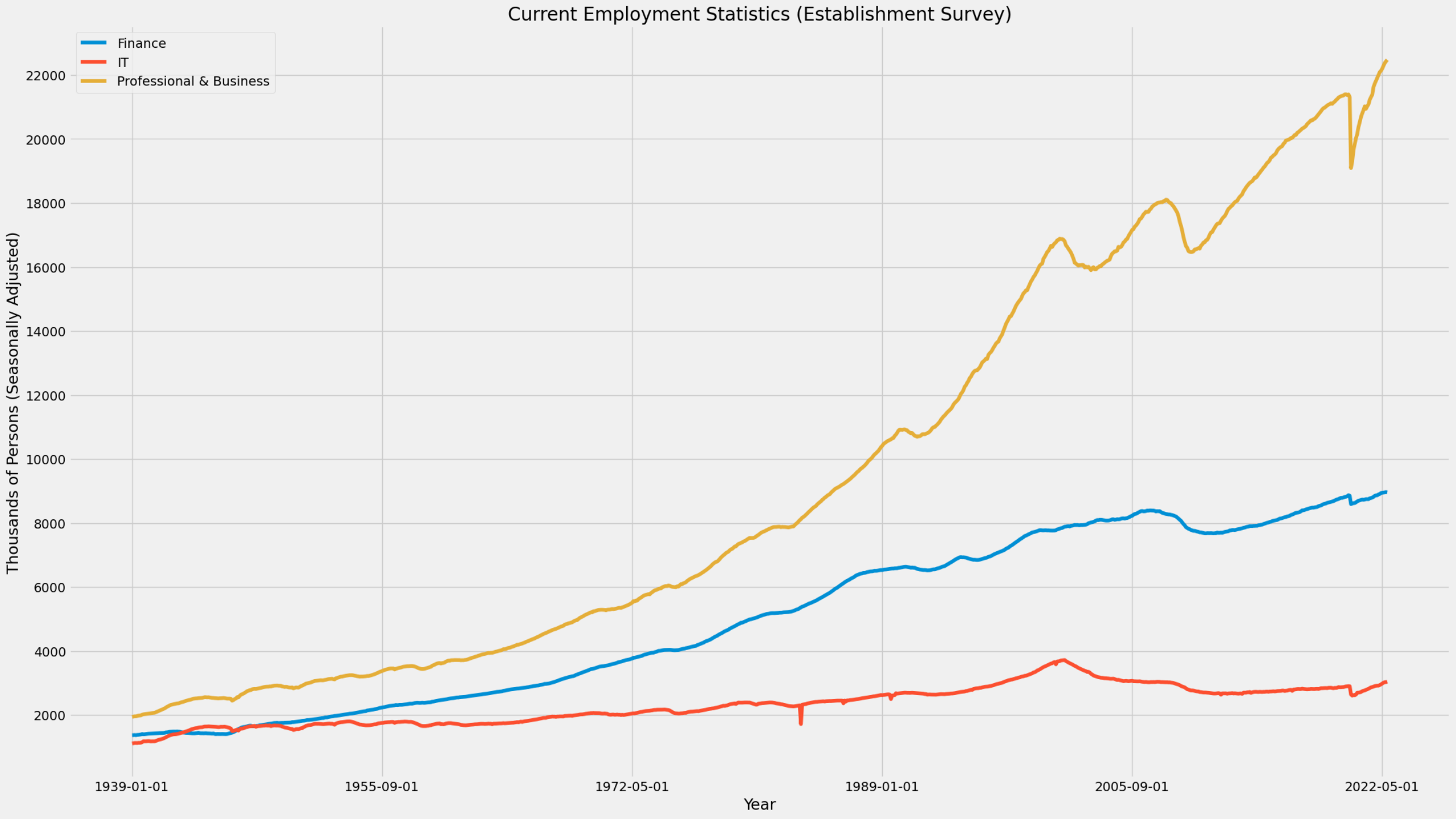






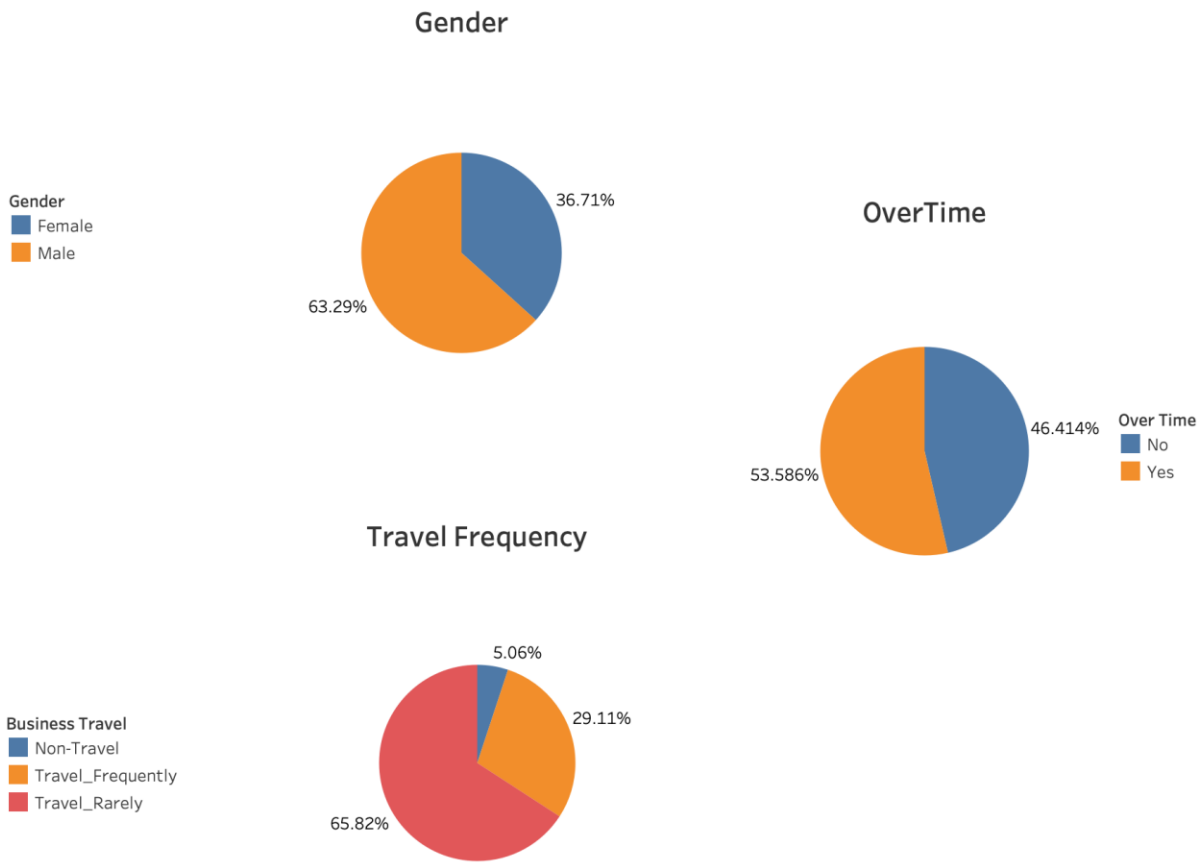
GEO-SPATIAL VISUALIZATION OF ATTRITION


Current Employment Statistics (Establishment Survey)
Manufacturing
Education and Health Services
Hospitality
25000
22500
20000
17500
15000
12500
10000
7500
5000
2500
Thousands of Persons (Seasonally Adjusted)
1939-01-01
1955-09-01
1972-05-01
1989-01-01
2005-09-01
2022-05-01
Year

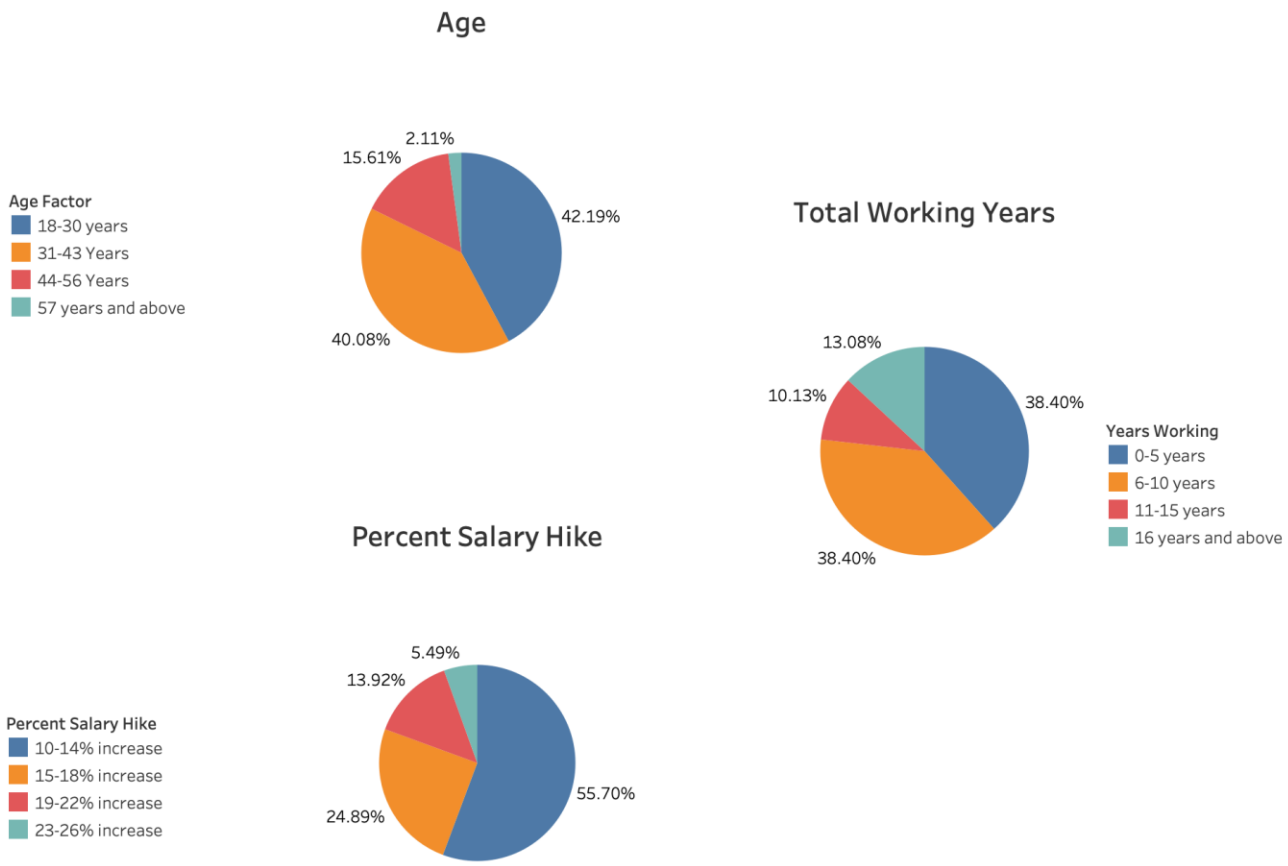

Current Employment Statistics (Establishment Survey)
Finance
IT
Professional & Business
Thousands of Persons (Seasonally Adjusted)
22000
20000
18000
16000
14000
12000
10000
8000
6000
4000
2000
1939-01-01
1955-09-01
1972-05-01
1989-01-01
2005-09-01
2022-05-01
Year

VISUALIZATION INSIGHTS - USING TABLEAU

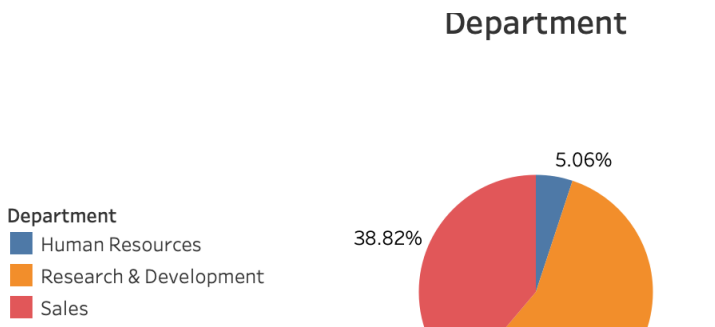
- Attrition Based on Gender
- Attrition Based on Over Time
- Attrition Based on Travel Frequency
- Attrition Based on Age
- Attrition Based on Total Working Hours
- Attrition Based on Percent Salary Hike
- Attrition Based on Department
- Attrition Based on Job Role
- Attrition Based on Marital Status

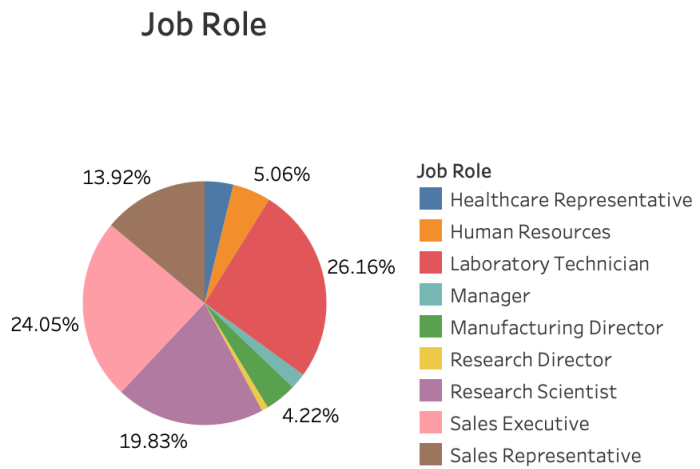


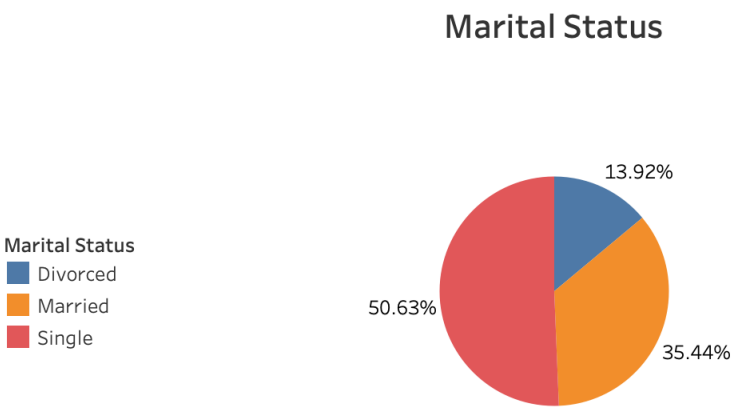
Attrition Based on



Attrition Based on







Attrition Based on

Thank You!