

# SAAFIR HALE

United States · (408) 337-7841

[saafirhale@gmail.com](mailto:saafirhale@gmail.com) · [saafirhale.github.io](https://saafirhale.github.io) · [github.com/SaafirHale](https://github.com/SaafirHale)

Passionate and dedicated Data Analyst with experience identifying efficiencies and problem areas within data streams while communicating needs for projects. Adept at receiving and monitoring data from multiple data streams, including Sheets, SQL, and Excel data sources. Skillful in various programming languages, including Python, R, JavaScript, C++, C#, Angular, Git, HTML, and CSS. Ability to synthesize quantitative information and interact effectively with colleagues and clients.

## TRAINING

**AUG 2020 – JAN 2021**

### KAGGLE

Achieved 10 Kaggle certifications to gain skills in Data Science. The Kaggle certifications completed were:

- Python
  - Explored functions, Booleans, lists, loops, dictionaries, and external libraries using Python
- Intro to Machine Learning
  - Learned about Data Exploration, model validation, and fine-tuning.
  - Built an ML Model that predicted Melbourne Housing Prices using a Decision Tree Regressor.
  - Gained a 23% decrease in validation error using a Random Forest Regressor.
- Intermediate Machine Learning
  - Utilized Python to replace missing values, encode categorical variables, and preprocess the data with pipelines.
  - Learned cross-validation to obtain an accurate assessment of model quality.
  - Added ensembles, like XGBoost, to achieve better performances.
- Data Visualization
  - Created line charts, bar charts, heatmaps, and scatter plots with the Python library Seaborn.
- Pandas
  - Extracted insights from a Wine Reviews Magazine using Pandas.
  - Grouped wine reviews by country and province and removed missing data from the table.
- Intro to SQL
  - Leveraged Google BigQuery and SQL to get insights, organize query results, and combine data from tables.
- Advanced SQL
  - Used JOINS and UNIONS to combine data from multiple tables.
  - Wrote efficient queries to run faster and use less data by utilizing nested queries.
  - Gained insights from using analytical calculations in SQL.
- Geospatial Analysis

- Identified regions and islands outside the non-profit Kiva's loan network, for them to find new Field Partners. Used GeoPandas to visualize the current Kiva network.
- Used a Coordinate Reference System (CRS) to plot the migratory path of the Purple Martin birds.
- Assessed which Japanese high-density prefectures are prone to high-magnitude earthquakes. Visualized the areas that will benefit from earthquake reinforcement using GeoPandas, Choropleth, and Folium.
- Used geocoding to repair missing latitude and longitude coordinates for Starbucks Berkeley locations. Selected counties where 100,000 households made at least \$150,000 per year, the median age is less than 38.5, and the density of inhabitants is at least 285 (per square kilometer), in addition also selected counties that meet one of the following criteria: at least 500,000 households making \$150,000 per year, the median age is less than 35.5, or density is at least 1400 (per square kilometer). The selected counties were used to consider a new Starbucks Roastery location in California.
- Utilized proximity analysis to identify the number of crashes that were outside the 10-minute range of New York City hospitals.
- Machine Learning Explainability
  - Learned the techniques needed to extract insights from sophisticated machine learning models.
  - Gained an understanding of how to *find* which features are important to an ML model using Permutation Importance.
  - Used Partial Dependence Plots to show *how* a feature affects predictions
  - Broke down a prediction showing *each* feature impact using SHAP Values (Shapley Additive exPlanations), SHAP Summary Plots, and SHAP Dependence Contribution Plots.
- Data Cleaning
  - Handled missing values in a dataset of issues San Francisco building permits by dropping columns and imputing the data.
  - Used Scaling and Normalization to transform numeric variables.
  - Parsed dates on a dataset containing information on earthquakes from 1965 to 2016.
  - Encoded characters to avoid Unicode Decode Errors.
  - Efficiently fixed inconsistent data entries.

## MARCH 2022

### SERVICENOW PLATFORM IMPLEMENTATION

- The course covered core system setup, configuration management database, LDAP, Single Sign-On, Web Services, REST API, Waterfall, Agile methodologies, end-to-end sprint cycles, and ServiceNow engagement processes.

## NOVEMBER 2022

### SERVICENOW DEVELOPER

- Used the ServiceNow GlideAjax API to create an application called Fetch. The Fetch app allowed pet adoption centers to automate the pet adoption request process.
- Created two tables, one for dogs to be adopted and the other for Adoption centers. Also created UI pages for both tables.

- Utilized JavaScript in ServiceNow Script Includes to create new dog records, send email notifications after UI action event, form user prompts for user information, and automate response follow-ups.

#### SEPTEMBER 2022

#### SERVICENOW ADMINISTRATOR

- Configured and imported databases on the ServiceNow Platform for Rotten Potatoes on Game Reviews.
- Customized the layout of the form and list tables of both Reviews and Games. Also created an Administration Group for Reviewers. Automated Game Review approvals with a custom workflow.
- Setup Business Rules, UI Policies, and UI Actions for Rotten Potatoes Reviews.

## PROJECTS

#### APRIL 2022

#### NHL OUTCOME PREDICTION MODEL, UNIVERSITY OF MICHIGAN SPORTS PERFORMANCE ANALYTICS SPECIALIZATION

- Used Pandas, NumPy, and SciKit-Learn to split the 2017 NHL data into training and test datasets.
- Built a prediction model using SciKit-Learn's GridSearchCV and DecisionTreeClassifier.
- The model created had an accuracy score of 53.1 % in predicting the game outcome of the 2017 NHL season.

#### APRIL 2022

#### BELLABEAT FITNESS TRACKER, GOOGLE DATA ANALYTICS PROFESSIONAL CERTIFICATE

- Imported hundreds of thousands of data to analyze the relationships from tracking data.
- Used R to clean, analyze trends, and design visualizations.
- Translated data into actionable questions, i.e. Do daily steps affect sedentary time?

#### MARCH 2022

#### CYCLISTIC BIKESHARE, GOOGLE DATA ANALYTICS PROFESSIONAL CERTIFICATE

- Analyzed 12 months of data to gain actionable insights on users.
- Utilized Excel and R for data cleaning and data visualization.
- Successfully discovered repetitive user habits and daily usage, for Cyclic subscription to shift bike share memberships.

#### FEBRUARY 2022

#### COVID-19 DATA EXPLORATION, PERSONAL PROJECT

- Collected millions of rows of global Covid data from a verified data source.
- Analyzed and formatted data in Microsoft SQL Server and Excel.
- Utilized Tableau to create a global Covid-19 dashboard.

## SKILLS

- Ability to Work in a Team
- Ability to Work Under Pressure
- Ability to Multitask
- Python
- Machine Learning
- JupyterLab/Notebook
- Pandas
- NumPy
- SciKit-Learn
- TensorFlow
- BeautifulSoup
- Seaborn
- Git
- JavaScript
- ServiceNow
- Saas
- R
- RStudio
- Ggplot2
- Tidyverse
- Dplyr
- SQL
- Excel
- Tableau
- PowerPoint
- HTML
- CSS
- Angular
- C/C++/C#
- Word
- Workflows

## EDUCATION

APRIL 2022

**GOOGLE DATA ANALYTICS  
PROFESSIONAL, GOOGLE**

APRIL 2022

**UNIVERSITY OF MICHIGAN SPORTS  
PERFORMANCE ANALYTICS,  
UNIVERSITY OF MICHIGAN SPORTS**

JUNE 2018

**HIGHSCHOOL DIPLOMA, PENN  
FOSTER HIGHSCHOOL**

Graduated early at the age of 15.

## ACTIVITIES

- Soccer
- Travel
- Reading
- Music

- Blockchain
- Game Development