

RELEVANT PROJECTS

- **Financial Data Science Project - Optiver Data:** develop a tool that predicts future volatility of financial data and communicate the tool using multimedia to a trader.
 - Conducted a comprehensive business analysis to define project objectives, followed by data collection, preprocessing, downsampling, and cross-validation to prepare datasets for analysis.
 - Developed and experimented with various time series models, including Holt-Winters, Simple Exponential Smoothing (SES), ARIMA, Time Series Linear Regression, Heston, Holt's Linear Trend Model, Autoregressive (AR), and ARIMA-GARCH.
 - Utilized evaluation metrics such as Mean Squared Error (MSE), Bayesian Information Criterion (BIC), running time, and model stability to assess and select the best-performing model.
 - Deployed the final model using a Shiny app in R Studio, creating an interactive tool to enhance real-time trading precision and decision-making support on the trading floor.
- **Optimal Portfolio Allocation for ASX Stocks:**
 - Develop a comprehensive skill set, including data analysis through calculating financial metrics like simple returns, statistical skills for deriving means and covariance matrices, and programming expertise with Python libraries like Pandas and NumPy.
 - Gain insights on portfolio optimization and risk management techniques, enhancing problem-solving abilities particularly in optimizing under constraints such as no short selling.
 - Sharpens ability to visually represent complex data through plots of efficient frontiers and risk-return profiles, and it improves decision-making under uncertainty by incorporating elements like risk-free assets into portfolio adjustments.
- **Predicting Age of Abalone: from Physical Measurements**
 - Spearheaded an age prediction analysis for abalone using RStudio, from initial problem identification through to data understanding and exploration. Investigated potential relationships within the data, employing correlation visuals to examine interactions between various variables.
 - Performed comprehensive data transformations to prepare for advanced modeling, utilizing techniques such as forward and backward selection and log-linear models.
 - Evaluated model performance rigorously using metrics such as RMSE, R-squared, and MAE, while also verifying assumptions and addressing multicollinearity to ensure robustness.
 - Developed and delivered insightful presentations and detailed reports, effectively communicating findings and recommendations to stakeholders through the strategic use of R-generated visualizations and narratives.
- **Greater Sydney Analysis**
 - Enhances geospatial data analysis through the use of tools like geopandas for GIS data manipulation, and advanced data processing skills in cleaning, transforming, and integrating datasets into SQL databases.
 - Refine statistical analysis techniques by computing and interpreting z-scores and their socio-economic implications. The project also develops your ability to create interactive maps and visual representations, essential for effective data communication.
 - Hones critical thinking through correlation analysis and report writing skills, culminating in a comprehensive presentation of findings. These competencies are crucial for careers in urban planning, data analytics, public policy, and GIS, making this a significant professional development experience.
- **Correlating Wealth with Health and Social Outcome: Comparative Analysis of Death Rates, Literacy, and Suicide Rates**

- Utilizes various regression techniques to understand the impact of socio-economic factors on health outcomes. Linear regression is employed to predict crude death rates from income and literacy levels, while K-Nearest Neighbors regression estimates these rates using suicide rates and income data. Bayesian Linear Regression forecasts a country's 2020 crude death rate based on data from 2010. Another linear regression model correlates crude death rates with suicide rates and income.
- The accuracy of each model is evaluated through coefficients, root mean square error, and R-squared values, culminating in a detailed visual comparison to assess the predictive effectiveness of different variables.