|  |  |  |  |
| --- | --- | --- | --- |
| **Project Number** | **Project Title** | **Statistical Technique** | **Author(s)** |
| 1 | Ergodicity Economics and Consumer Preferences Under Additive and Compound Wealth Growth | Model Selection | Himesh Buch |
| 2 | Predicting a Pitcher’s Performance with Linear Regression | Simple Linear Regression (1 variable) | Nadine Burgos, Sheng Xi Chen |
| 3 | Mixed Procedure and Covariances | Mixed Model: Random Intercept, random intercept, random slope | Binghong Cai |
| 4 | Cross Validation on Graphs | Perform 10-fold cross validation on baseline machine learning models | Advith Chegu |
| 5 | Time Series Forecasting | Forecasting using monthly car sales in Quebec 1960-1968. Included seasonality. | Abdallah Dar |
| 6 | Multiple Linear Regression: Live Expectancy Between the States Model | Model selection. Predicting the average life expectancy of the habitants of different states. | Nicole Del Villar |
| 7 | A Demonstration of Model Selection using Stepwise Regression and K-Fold Cross Validation | Model selection. Demonstrate model selection on Happiness scores of all countries. | Samir Fidai |
| 8 | Simple Forecasting Models Using R | Unemployment rate from  Jan. 2010 to Dec. 2019. Naïve Method, Drift Method, Average Method, and Seasonal Naïve Method. | Samantha Gong |
| 9 | Using Monte Carlo Simulations to Estimate Pi | Monte Carlo simulations were performed. | Sahil Gurugunti, Deivanayagam Shanmugasundaram, Aashneil Uppuluri, Neha Agarwal |
| 10 | Forecasting Stock Price Using Time Series Analysis | Used three methods for forcasting: Simple Moving Average, Exponential Smoothing  ARIMA Model. | Jiali Han, Junhua Zhu, Wei Li, Zhuofan Dong |
| 11 | Model Selection & Cross-Validation for Marijuana Arrest Data | Forward, backward, bi-directional Model Selection. Arrests for marijuana possession. | Anna Atkuru, Rachel Altschuler, Alexandra Jacko, Celine LaBelle |
| 12 | MODEL SELECTION AND CROSS VALIDATION: Licks to the Center of a Tootsie Pop | Contributing factors to average licks were analyzed via model selection and cross validation. | Marissa Jensen |
| 13 | ANALYSIS OF THE MOST RECENT 399 SONGS IN MY SPOTIFY | Exploratory analyses including graphical methods for variables associated with songs. | Andrew Jofre |
| 14 | DETECTING UNUSUAL OBESERVATIONS | Outlier detection and outlier impact. | Yeoun Chan Kim |
| 15 | Data Analysis Plots | Scatter, Stem-and-Leaf, Histogram, Box and Whisker, Ellipse, Residual, Quantile-Quantile graphical displays. | Varun Krishnan, Shreya Patel, Abhishek Modoor |
| 16 | Weight loss and Self-esteem | Used a linear model approach to compare 3 different diet methods. | Hyun Jin Ku |
| 17 | Outliers and Trimming | Searched for outliers and effective ways to deal with them. | Emily Li |
| 18 | Model Selection - Different Validation Approaches | Performed validations using cross-validation and bootstrapping to find the optimal model for house pricing data. | Xinjie Li, Baoluo Sun |
| 19 | U.S. Insurance Policy | Performed model selection to predict the insurance charges of a major U.S insurance company. | Guangjin Luo |
| 20 | Model Selection in R for MLB Wins as a Function of Team Batting Statistics | Used regression model selection techniques to analyze potential models to forecast wins in major league baseball. | Thomas Maguire |
| 21 | Statistical Analysis of Pima Indians using R | Performed analyses to understand the variables in the Pima data set in order to build a stepwise regression model for predicting diabetes. | Arpan Niyogi |
| 22 | Can you predict an NFL player’s success based off of their combine performance? | Used classification trees to predict the success of an NFL player based on their NFL combine statistics. | Sean O'Sullivan |
| 23 | Linear Regression in R | Built a linear regression model to predict sales. | Dhruv Shilotri |
| 24 | K-Means Clustering of Forest Types Using R | Used k-means clustering based on spectral information representing forest types to classify data into four possible forest types. | Eugenie Shin |
| 25 | Statistical approaches to determine and assess the quality of different types of wines | Use multivariable statistical analyses to determine and assess the quality of different types of wines. | Tiansheng Tan, Qingyuan Xu, Howard Young |
| 26 | Model Selection for Vehicle MPG | Performed feature selection followed by model selection to determine the best model to predict mpg in vehicles. | Preston Tolbert |
| 27 | Linear Regression Analysis: The Importance of Validating Your Linear Model | Used stepwise procedures to predict how much weight a male should be able to bench press. Provided Power Point slides and a Video explaining the methodology used for model building and validation. | Dimitri Victor |
| 28 | Passenger Prediction Survival in Titanic Using R | Used Titanic data to predict survival of passengers on the Titanic. | Jinnian Shen |
| 29 | Model Selection in R for College Graduate Earnings | Analyzed financial data on college graduates to conduct stepwise selection and k-fold cross validation to predict earnings and debt. | Yimeng Wang |
| 30 | Monte Carlo Simulation of American Roulette in R | Created a roulette wheel simulation using R to model starting bet strategy. | Matthew Wechsler |
| 31 | R in Animal Crossing | Used graphs and line charts to detect patterns in turnip price to determine the best price to sell turnips in the video game “Animal Crossing.” | Jinhan Huang, Ailin Zhang, Keren Zhang |
| 32 | Model Selection in R for Ozone Levels in Los Angeles 1976 | Model selection with cross validation to predict using ozone concentration data. | Yuqing Zhou, Mazin Rafi |
| 33 | Factors influencing the numbers of COVID-19 infected people in the United States | Utilized stepwise and cross-validation methods to build the best model to roughly estimate the number of COVID-19 infected people using population related factors. | Yuting Zhu, Jianze Wang, Heyuan Huang |
| 34 | Resampling Method in R | Develop a resampling scheme when comparing two groups in an experiment to obtain a p-value. | Che Zhang, Ge Qu |
| 35 | The House Always Wins | Monte Carlo simulation of Blackjack | Jake Seary |
| 36 | Boston Housing Analysis | Used box plots and tree plots to explore the relationship between demographics and crime. | Rutul Tijoriwala |