

# Benedict Neo

☎ 515-388-0996 ✉ benedict.neo@outlook.com 🔗 linkedin.com/in/benedictneo 🐙 github.com/benthecoder

## Education

### Iowa State University

Expected: Dec. 2023

Bachelor of Science, Major: Statistics, Minor: Computer Science

Ames, IA

- Cumulative GPA: 4.0
- Coursework: Experimental Design, Probability and Statistics I & II, Data Structures & Algorithms, Discrete Math

## Experience

### Tesla

May 2022 – Aug. 2022

Data Science Intern

Fremont, CA

- Developed and revamped 10 PowerBI dashboards tracking KPIs for various operations and systems within the factory
- Reduced PowerBI dashboard data refresh time by **80%** by rebuilding the data model and writing efficient DAX queries
- Utilized **NLP** techniques on work orders which led to creation of failure codes enabling better insights for asset failures
- Investigated work orders and identified assets where **\$100,000** per year can be saved in preventive maintenance costs
- Cleaned and analyzed shift hours and badging data to provide insights into optimizing parking space during peak hours
- Detected anomalies and trends by performing time series analysis for the cooling tower systems at the factory

### Iowa State University

Jan. 2022 – May 2022

Undergraduate Research Assistant

Ames, IA

- Developed and contributed to **WEPPR**, an R package that emulates the Water Erosion Prediction Project model
- Wrote functional code in R (Tidyverse) to clean and transform 100GB of raw soil and climate data into tidy format
- Authored clear documentation for functions and classes in the package with the roxygen2 package
- Developed unit and functional tests to ensure proper functioning of code with the testthat package

### bitgrit

Aug. 2021 – Dec. 2021

Data Science Intern

Remote

- Led data collection and cleaning, and authored end-to-end solutions in Python for data science competitions
- Managed the company data science publication and published 30+ technical articles with over 500,000 views [🔗](#)
- Taught 50+ people in a workshop best practices and common tasks for data cleaning in Python

## Projects

### 🐙 GPT-3 Blog Title Optimizer | Python, Selenium, BeautifulSoup, OpenAI, Weights& Biases, BigQuery, SQL, Streamlit

- Scraped 200k blog articles from medium.com with Selenium and BeautifulSoup and ingested data into BigQuery
- Fine-tuned transformer models with OpenAI API and tracked experiments and datasets with Weights & Biases
- Containerized the application with Docker and deployed the model as a web app using Streamlit for users

### 🐙 Energytics | Python, Streamlit, Pandas, Plotly, Scikit-learn, LightGBM, OpenWeatherMap API

- Built a web app that presents insights on energy production cost and building energy usage in the US
- Conducted EDA on 60M records with 18 features to investigate trends, outliers, missing data, and anomalies
- Trained and deployed a LightGBM model to predict building energy usage from user location and building details

### 🐙 LinkedIn Insights | Python, Streamlit, Pandas, thefuzz, Plotly, Pyvis, Networkx

- Developed a web app with 500+ users that allows them to gain insights into their LinkedIn connections
- Utilized fuzzy matching to clean and manipulate raw user data for more accurate insights
- Visualized user connections with interactive bar charts, time series plots, and network graphs

## Skills

**Languages:** Python, R, SQL, SAS, JavaScript (React), Java, HTML/CSS, Bash

**Libraries:** Pandas, NumPy, Matplotlib, Plotly, Tidyverse, ggplot2, Scikit-Learn, LightGBM, NLTK

**Tools:** AWS, Google Cloud, Docker, Power BI, Tableau, Looker, Git, Linux, MongoDB, Excel

**Data Science:** Data cleaning, Data visualization, Hypothesis Testing, Regression, Classification, NLP

## Leadership

### President of Google Developer Student Club

Aug. 2022 - Present

Iowa State University

Ames, IA

- Led a team of 12 core team officers to organize and host tech workshops and talks to over 100+ students