# XINYU HUANG

70W93, Apt. 6D
New York, NY 10025

917-455-6698
Xh2511@columbia.edu

## EDUCATION

**Columbia University** New York, US
**Master of Science in Data Science, GPA: 4.08/4.0** Expected Dec 2022

Coursework: Computer Systems, Machine Learning, Data Structure and Algorithms, Probability and Stats, Data Analysis and Visualization.

**Zhejiang University** Zhejiang, CN
**Bachelor of Science in Mathematics, GPA: 3.88/4.0** Jun 2021

Coursework: Data Structures, Programming in R, C++, Python, Optimization Algorithms, Multivariate Statistic Analysis.

## RESEARCH EXPERIENCE

**Multi-view Image Imputation with Adversarial Optimal Transport Networks** Zhejiang, CN
**Research Assistant. Advisor: Prof. Xiaoye Miao** Jul 2020 - Apr 2021

- Imputed missing part of images using GAN architecture with python packages Tensorflow, Opencv and Numpy.
- Developed a multi-view encoder model to build coordinate representation of embeddings from five different views.
- Built a decoder to reconstructed images and optimized whole model minimizing two losses including reconstruction loss and adversarial optimal transport loss.
- Trained a discriminator jointly with generator to distinguish difference between real data and generated data to assist training of generator.
- Achieved better performance on four real-life datasets compared to five baselines.

**A Survey on Modern Data Pricing Models** Zhejiang, CN
**Research Assistant. Advisor: Prof. Yunjun Gao** May 2019 - Nov 2019

- Conducted a comprehensive survey on modern data pricing solutions based on dozens of articles.
- Introduced concepts of data markets and five existing commercial data markets.
- Categorized data pricing solutions into five pricing models and listed pros and cons: database based pricing model, ML based pricing model, game theory based pricing model, auction based pricing model and pricing in social networks.
- Investigated how different pricing solutions are applied and maximize overall utilities in three real-life application scenarios: pricing in mobile crowd sensing, pricing in real-time ridesharing, online advertisement auction.

## DATA SCIENCE PROJECTS

**2021 Columbia Data Visualization Course Project, Tenants' Guidebook to New York City** Sep 2021 - Dec 2021

- Led a team of two to evaluate overall living conditions including COVID-19 situations, crime rates and air qualities across all neighborhoods in New York City and offer some suggestions on finding a safe and healthy neighborhood.
- Gathered and cleansed data all over NYC and utilized heat-maps, box-plots and time series to visualize how these three types of data distributed.
- Constructed an interactive component with html to display how air quality changed in last decades within different neighborhoods.
- Obtained conclusions based on analysis and provided four helpful neighborhood suggestions for different needs.

**2021 Columbia Data Science Hackathon, Air Quality Analysis and ML based Data Prediction** Oct 2021 - Oct 2021

- Led a team of three to analyze air quality data in US and make use of multiple machine learning skills to further explore data to predict future air conditions.
- Integrated and cleaned four air quality datasets from all around country in last decades with pandas for further applications.
- Interpreted and visualized air quality data and obtain three conclusions on air quality changes.
- Utilized three ML skills linear regression, ridge regression with kernel trick and deep neural networks with past air quality data to predict air conditions in 2021. Achieved a high out-of-sample accuracy.

## SKILLS

- Google Cloud, Python, SQL, R, C++, Java, MATLAB, Latex.
- Numpy, Pandas, Py-torch, Tensorflow, Opencv, Matplotlib, Ggplot2.
- Mathematical analysis, Statistical analysis, Optimization.