

Data Analysis Final Assignment

Andreas Heindl

I Love Data Analysis

The Data Solar Power Plant

Dataset from Kaggle

15-min time series over 2 months

~70k datapoints

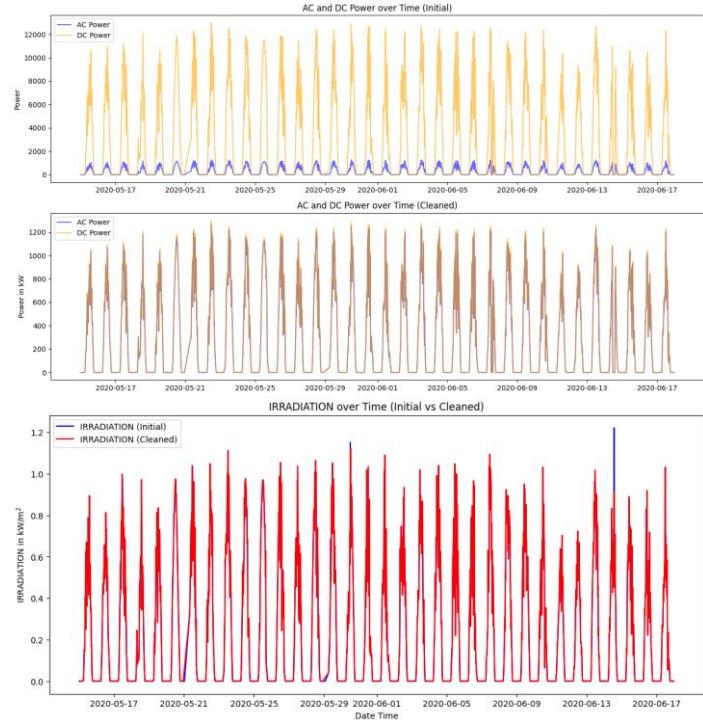
Data cleaning & Preprocessing

DC Power fix

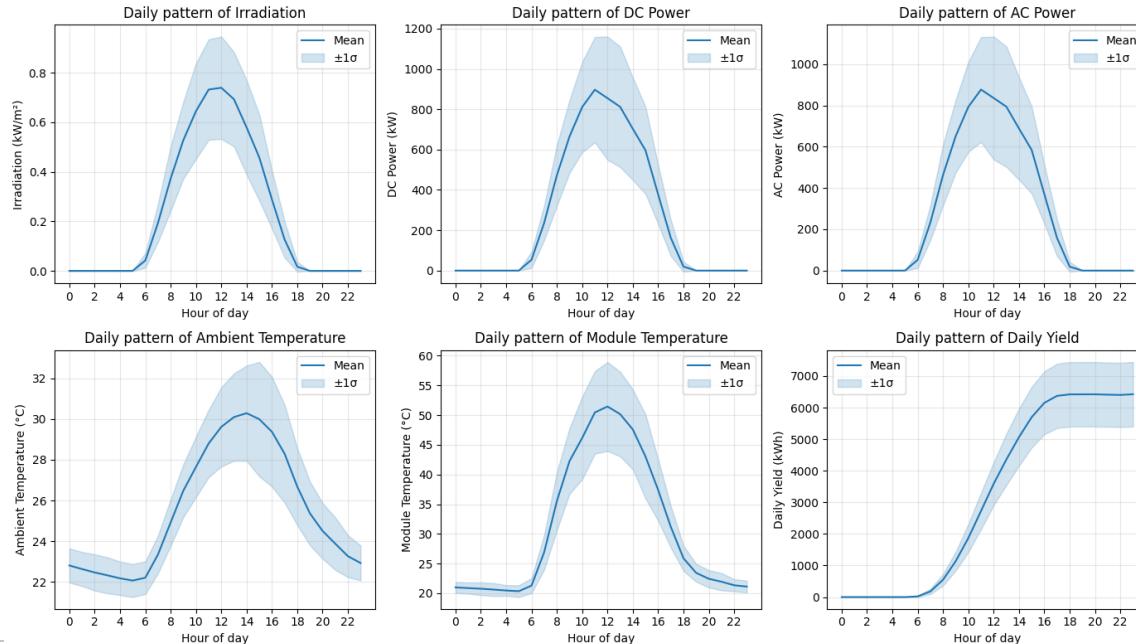
Handling missing and outliers

2 different files merged:

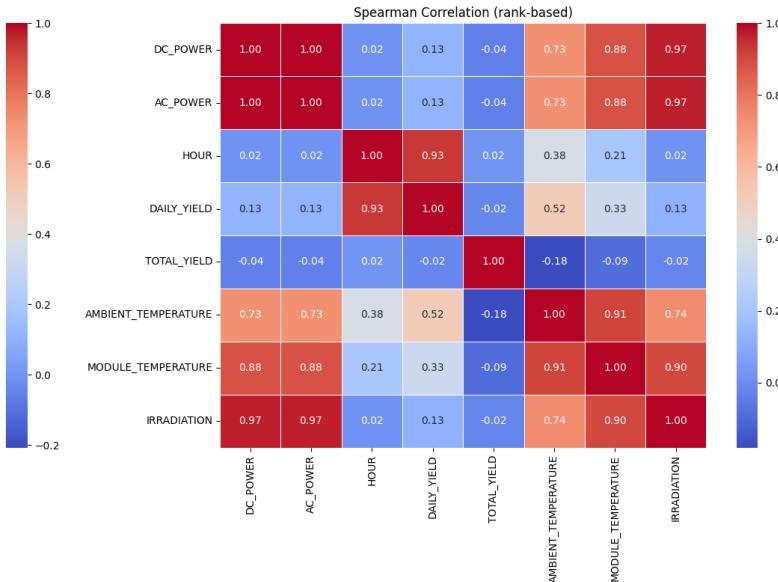
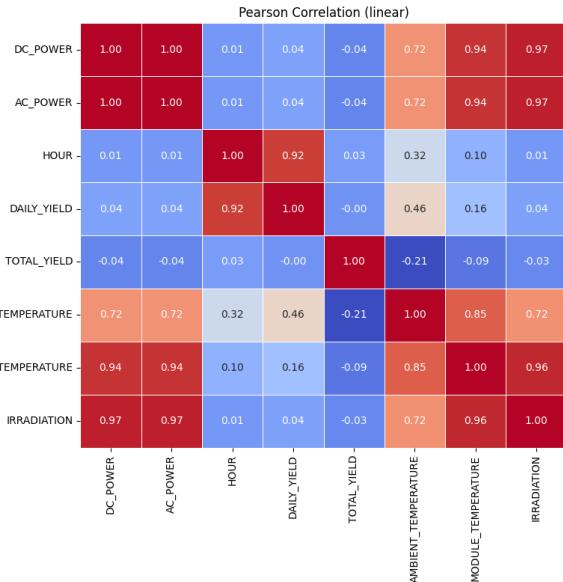
Weather and Generation



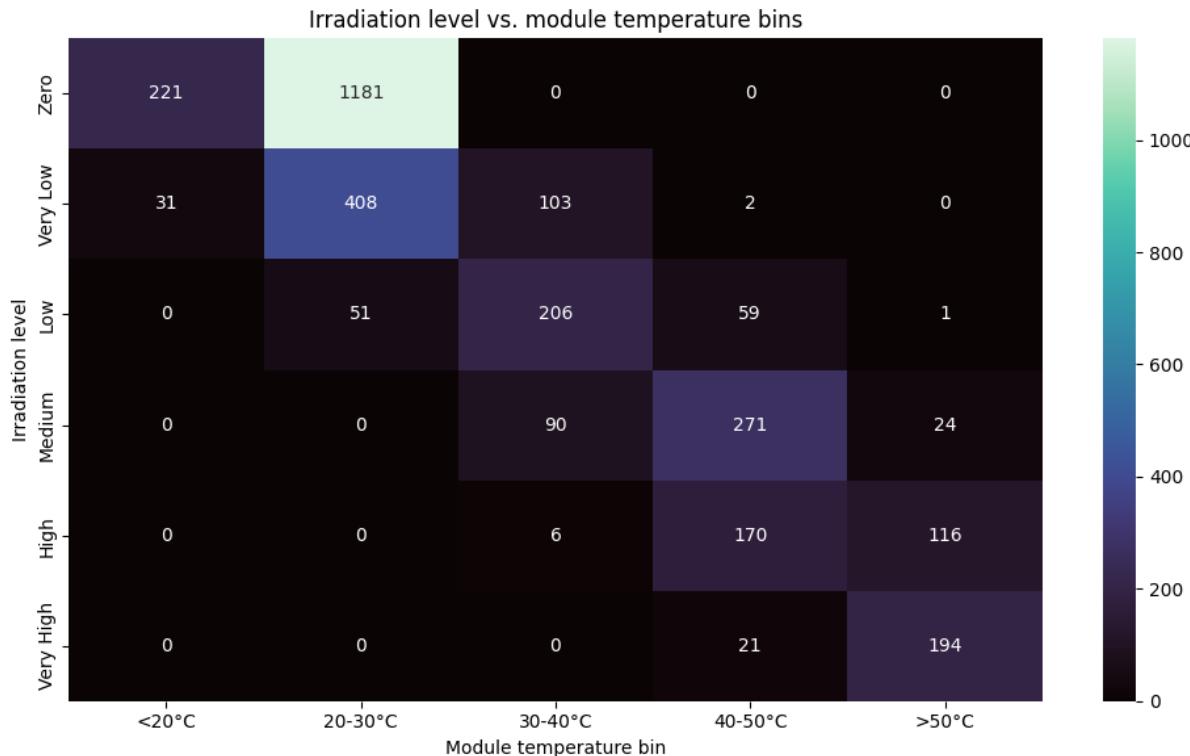
Day and Night cycle



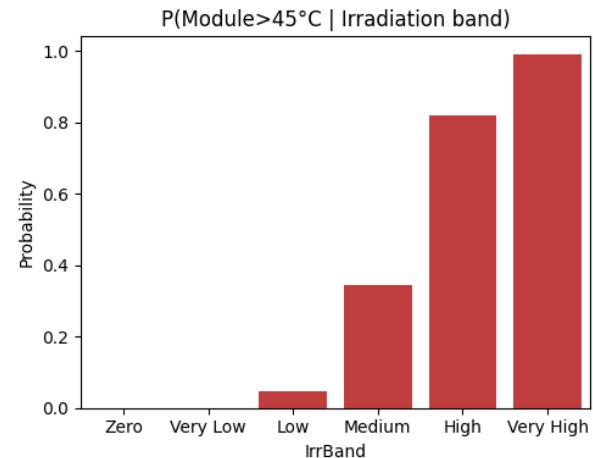
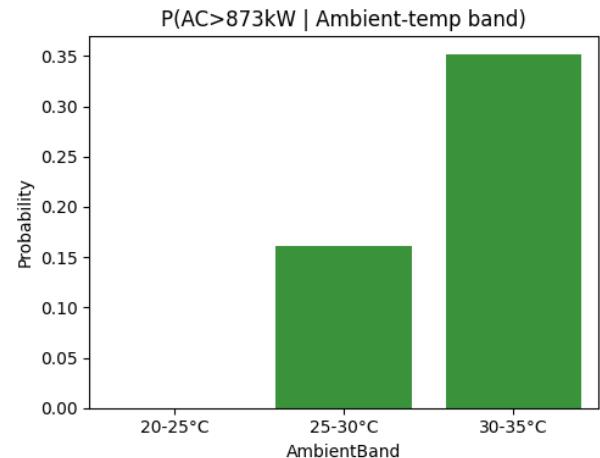
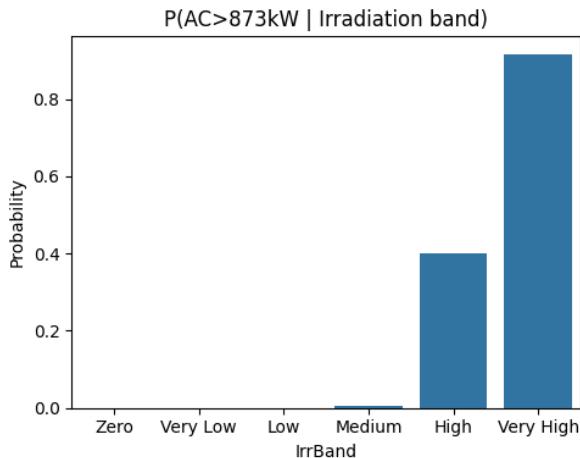
Correlation Analysis



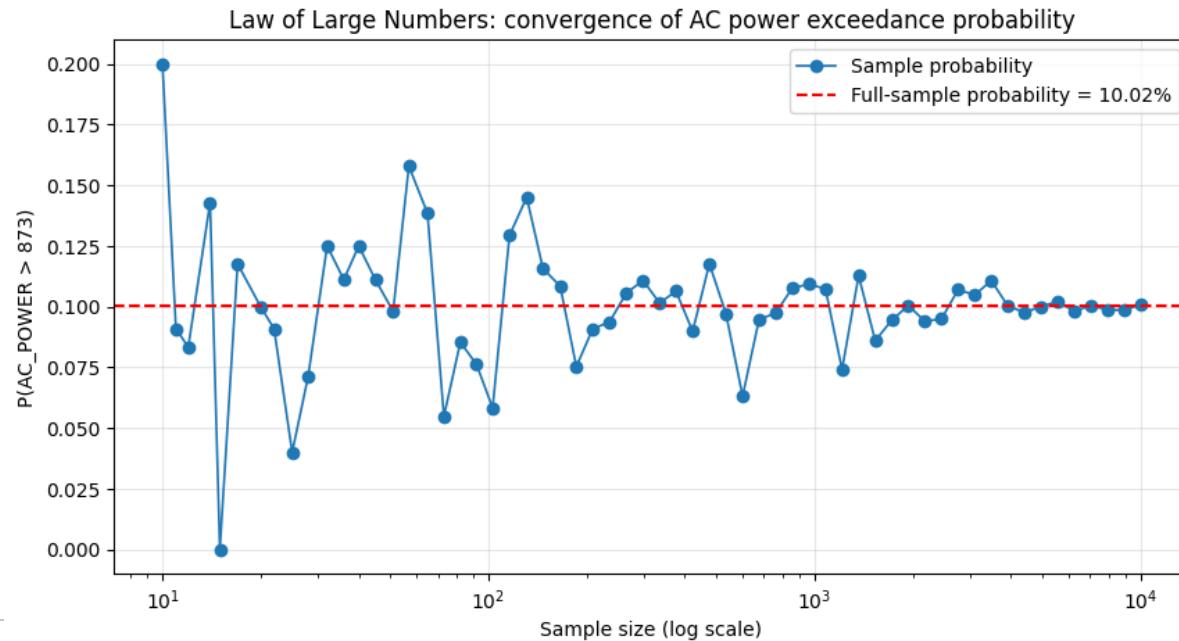
Cross Tab



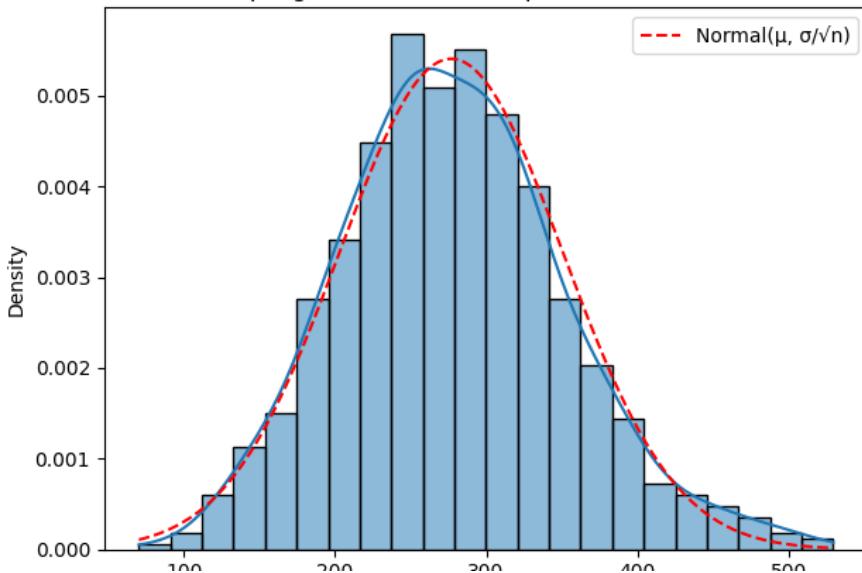
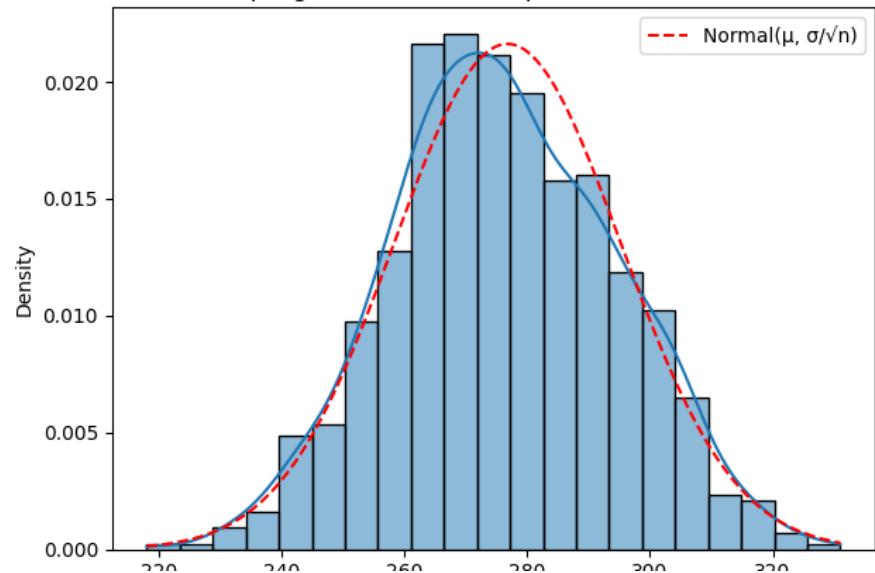
Conditional Probability



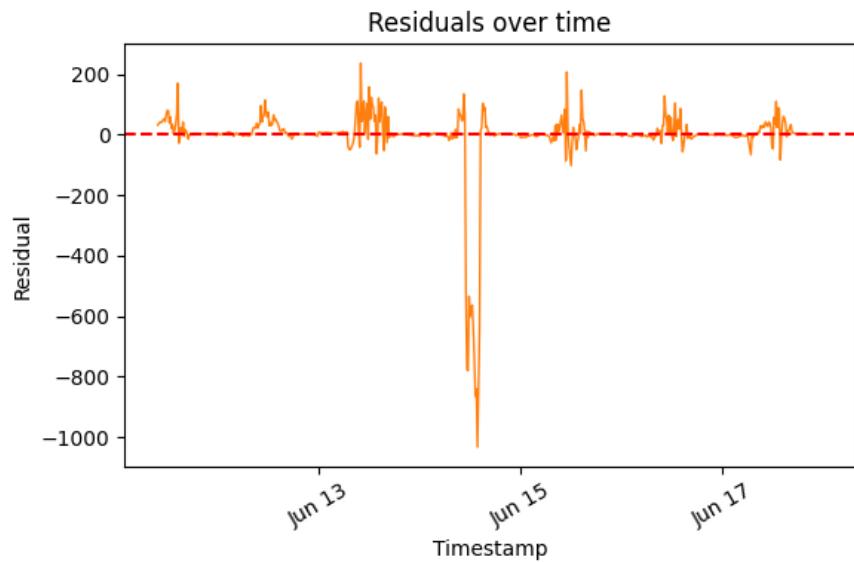
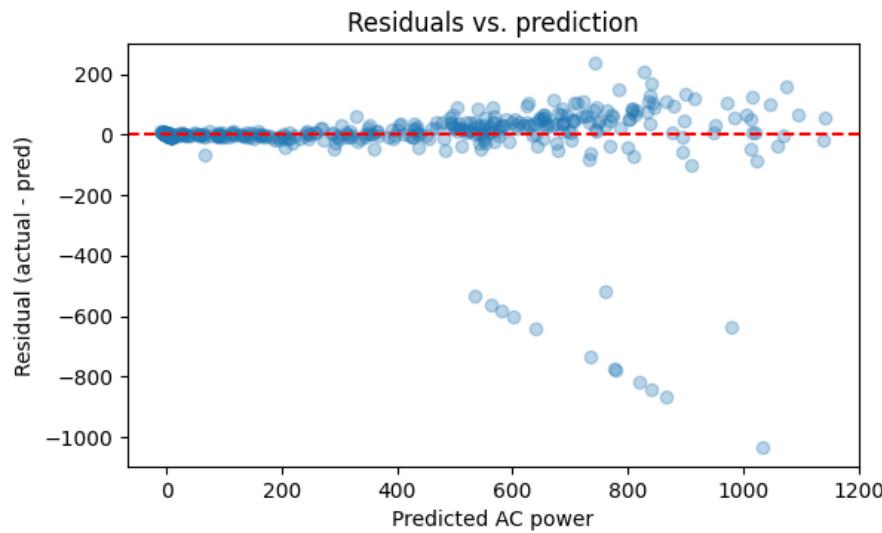
Law of Large Numbers



Central Limit Theorem

Sampling distribution of AC power means ($n=24$)Sampling distribution of AC power means ($n=384$)

Regression and Predictive Modeling



PCA, t-SNE, UMAP

