

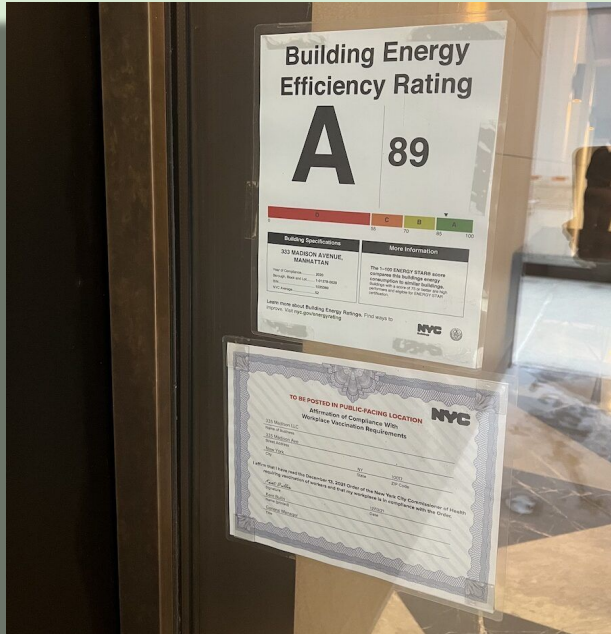


Energy Star Score

Measuring Building Efficiency in NYC

Jahnavi Kalpathy, Data Scientist
February 6, 2023

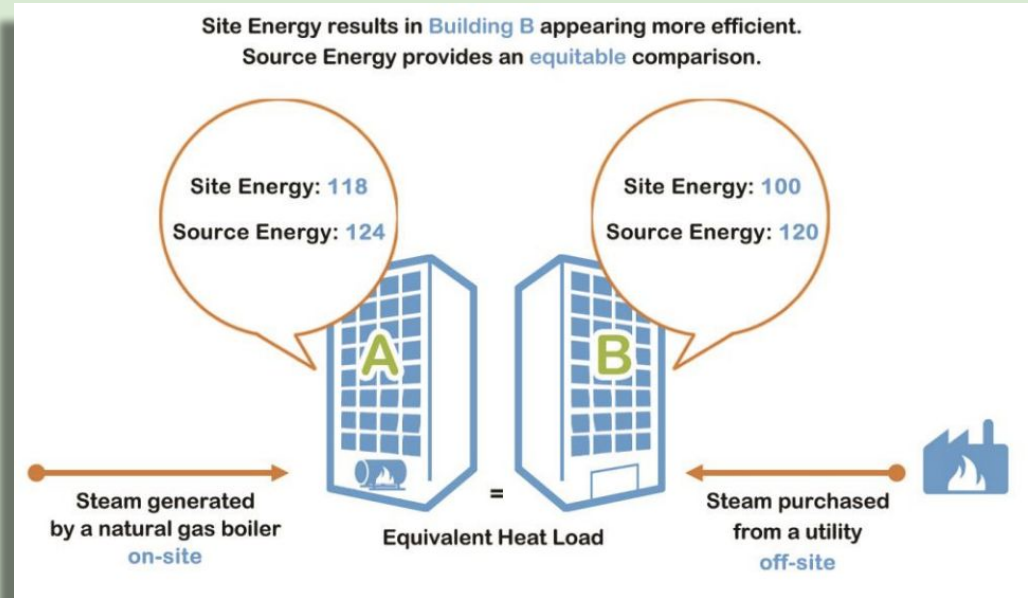
Portfolio Manager



How the Score is Calculated

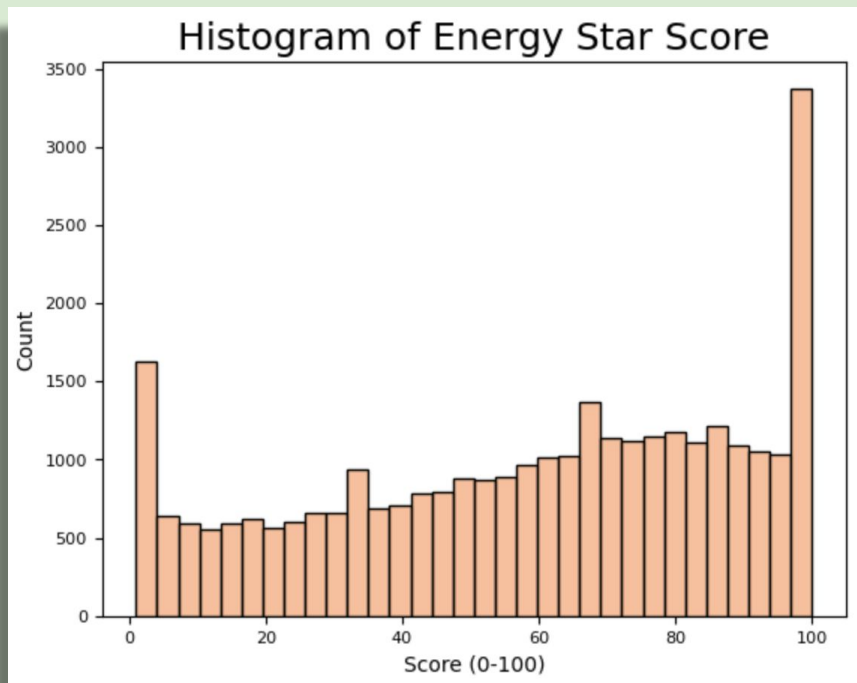
Question:

- what are the limitations of the Portfolio Manager calculation?
- Are there specific features that tell us more about the building and its score?



From Portfolio Manager Technical Report

The Energy Star Score



Dataset covered over 40,000 observations, which represent individual buildings (or campuses) across greater New York City.

Unfortunately not a normal distribution
→ possible to transform the target?

Data Collected per Building

Utility Bill

- Annual energy consumption
- Annual water consumption

Smart Meters

- Peak consumption
- Consumption by area

Building Info

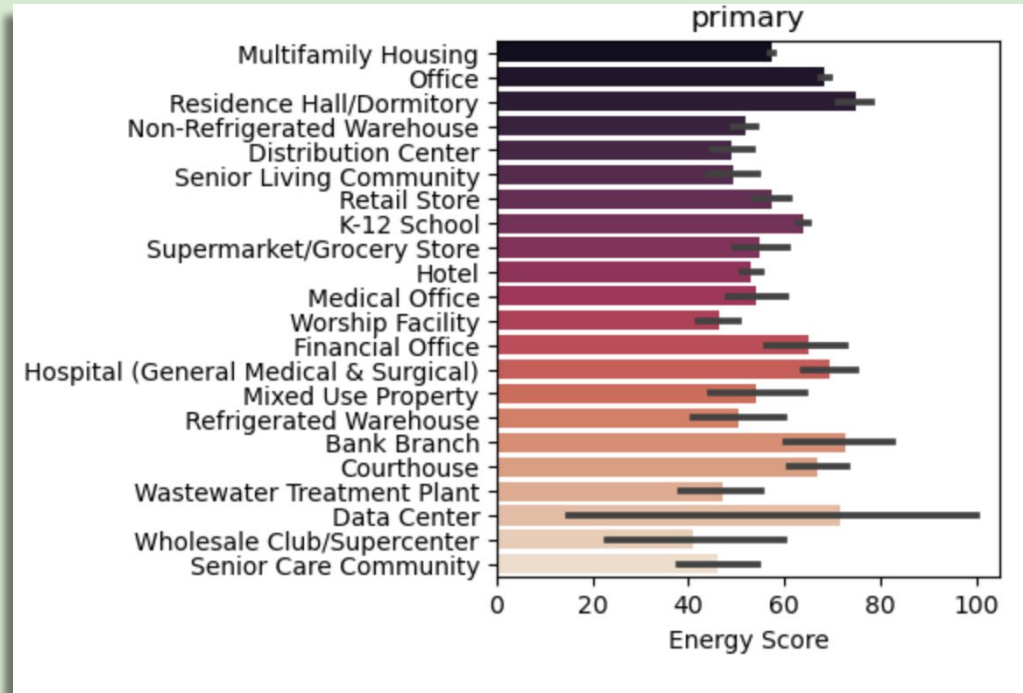
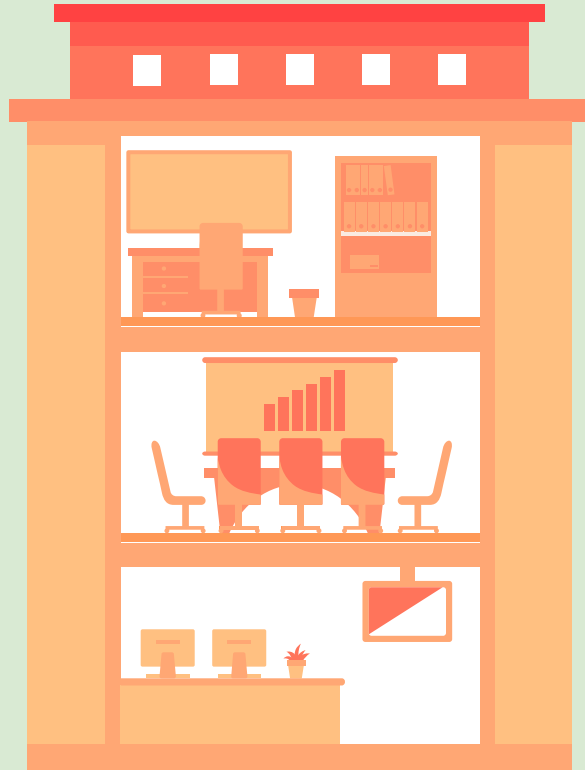
- Occupancy
- Construction Status
- Primary Property Use

Location Info

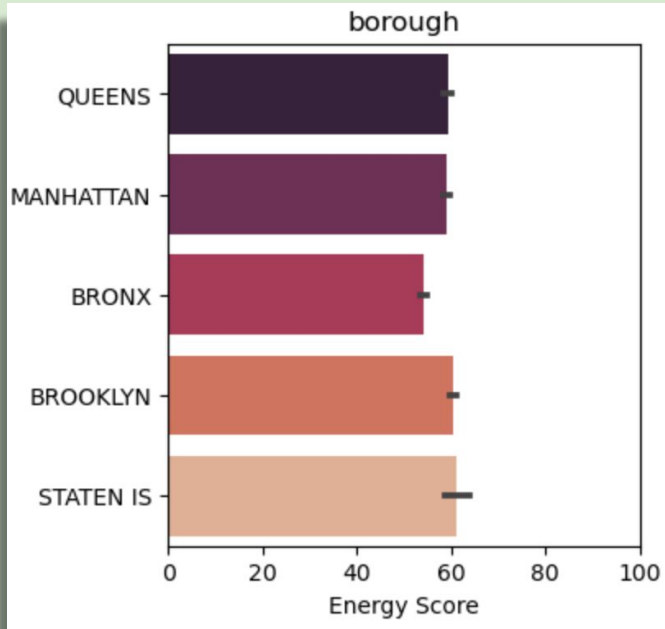
- Address
- Borough
- Weather Data (separate)



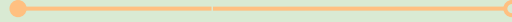
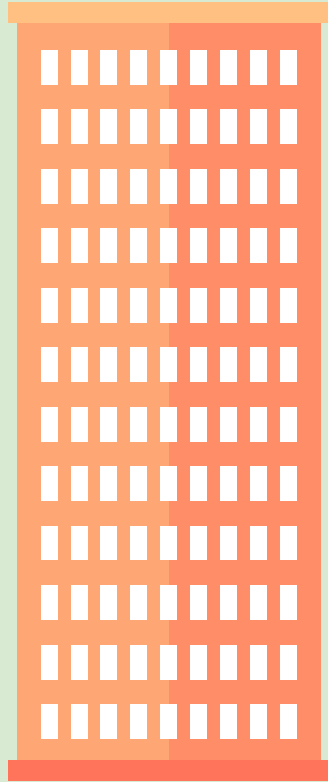
Primary Property Uses



Boroughs

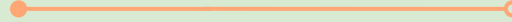


Feature Engineering & Preprocessing



Imputation

Removing features with more than 50% nulls and imputing rest



One Hot Encoding

Encoding `property_type` and `borough`



Polynomial Features

Feature interactions for various energy and water measurements



Standard Scaling

Scaling features around 0

Modeling

	cross_val_r2	train_r2	test_r2	train_rmse	test_rmse
model					
baseline	NaN	NaN	-0.000	30.0	30.0
LinReg-pf	-8.802112e+09	0.789	-3067.251	14.0	1662.0
Ridge-pf	3.880000e-01	0.467	0.435	22.0	23.0
Lasso-pf	3.610000e-01	0.489	0.432	21.0	23.0
ElasticNet-pf	4.200000e-01	0.515	0.450	21.0	22.0



Conclusions

- Too many of the features are enmeshed
- Portfolio manager calculates Energy Star Score AND other values like Total Emissions
- Need additional data related to weather normalization/property type normalization



Next Steps

- Additional attempts with target transformation
- Change target of prediction
 - Predict property type based on energy data?
 - Create clusters to identify new correlation (type of funding?), might need additional data
- Incorporate additional data



THANKS!

Do you have any questions?

CREDITS: This presentation template was created by **Slidesgo**, including icons by **Flaticon**, and infographics & images by **Freepik**