# Global

## Assumptions

* Time is the only variable necessary for projecting XYZ. Justification: Projecting XYZ requires finding XYZ as a function of time. In reality XYZ may be affected by an indefinite number of variables. Under our assumption, each of these factors is assumed to depend on time. Such that, the sum of the time derivatives of each factor is equal to the time derivative of XYZ. Hence observing the trend of XYZ can describe the summed effects of each factor, negating any need to consider individual factors when developing the model.
* UNKNOWN WILL BE INSIGNIFICANT IN LONG TIMESCALE DUE TO REDUCING TREND SO WE CAN IGNORE THE UNKNOWN HOUSES
* Data from the deep past can be said to be irrelevant.
  + Technological advancements make it so.
* The government will not create additional legislation on XYZ in the near future of our models. Justification: Such regulations would have a large impact on XYZ. For our models to be consistent, these changes cannot be taken into account.
* Staggering data values by year is negligible. Although causal effects typically take some lag time to manifest, all variables most likely cause an increase in e-bike sales within the same year the variable changes. Since our data is discretized by year, the time lag will be trivial.

## Model Development

## Advantages

## Disadvantages

# Linear

## Assumptions

Population will stabilize => household number will stabilize. There exists a carrying capacity of household numbers.

* Assumption: QUANTITY will not drop below critical value. Justification: QUANTITY has had a consistent downward trend, it is bounded because the percent cannot drop below 0. Even as QUANTITY dwindles, it is reasonable that the use will stabilize at some low percentage because JUSTIFICATION. We approximate the stabilization point to be around CRITICAL VALUE.
  + We refer to our assumption that QUANTITY use will not drop belowCIRITCAL VALUE and apply a lower bound to model

the choice of a linear approximation can be justified by simply noting the local linearity of the function over the relatively short time

span examined, which was especially appropriate in light of the small changes seen from year to year.

## Model Development

* We found the r value to be X, which suggests a weak/strong negative/positive linear relationship. Also R^2 = Y%, so Y% of the variability in the quantity can be explained by the linear model relating the proportion of use to the time in years.

## Advantages

## Disadvantages

# Exponential

## Assumptions

* Include a comment on long-term suitability of model example:
  + model is accurate up until the short-term
    - because changes in the exponent coefficients cause minimal changes (minimal error) in XYZ in the short run
    - Unpredictable so any radical market changes can not be added to our model, this is the best we can do
* Subdivide population and make assumptions on each case - this adds depth to our answer
* https://www.investopedia.com/terms/h/homoskedastic.asp#:~:text=Homoskedastic%20(also%20spelled%20%22homoscedastic%22,of%20the%20predictor%20variable%20changes.

Homoscedasticity and heteroscedasticity

Residuals on the overall linear plot is NON constant, implying heteroskedasticity, which suggests there are underlying factors at play, and there could be cross talk present between these factors (similar to multicollinearity).

A more thorough logistic regression for each type of household could specify a yield a more accurate model

The first order sobol index wrt. a feature can ascribe importance to each feature, allowing us to forecast which of the features contributes the most to the housing supply in time.

## Model Development

* Justify exponential - can use laws e.g. Moore’s law
* Uses Scipy’s curve fit function - Least squares
* Explain zero asymptotes if present

## Advantages

* simplicity and conciseness of the model

## Disadvantages

* As such, we do not account for the carrying capacity and growth cap. Accordingly, our model should not be used to project extensively far into the future.
* model is accurate only in the short-term because changes in the exponent coefficients cause minimal changes (minimal error) in XYZ in the short run

# Logistic

## Assumptions

No robust way to calculate carrying capacity in considered timescale. We chose to do UK because the data is more available.

## Model Development

* Our team believed that the best model for this spread would be a logistic model due to its two main characteristics: exponential growth and a limit at some fixed capacity.
* L is carrying capacity
* Calculate a value for carrying cap, do +- 15%

## Advantages

## Disadvantages

# Meyer Notes

* Uses Python Scipy libraries curve fit function - uses least squares (explain)
* Firstly code data to make first year (e.g. 2006) 0
* Logistic uses formula L/(1 + e^-k(x-x0) - search up meaning of each and explain
* We need to calculate L