An Introduction to Impact Charts

Darren Vengroff, Ph.D.
vengroff@datapinions.com

Agenda

I. Background
II. What is the problem?
III. Impact chart methodology
IV. Validation with synthetic data
V. How to generate an impact chart in three lines of code
VI. An application with real data

Background

Related Work: Prediction and Explanation

- Machine learning (ML) tends to focus on prediction
  - What video will a person watch?
  - Will a person pay back a loan?
  - What word comes next?
- Statistical social sciences tend to focus on explanation
  - Does going to college lead to higher income?
  - Are Black people discriminated against in access to credit?
  - Causal models are gold standard; regression methods widely used
- Interpretable machine learning
  - Are ML systems biased?
  - Can we explain why an ML system made certain predictions?

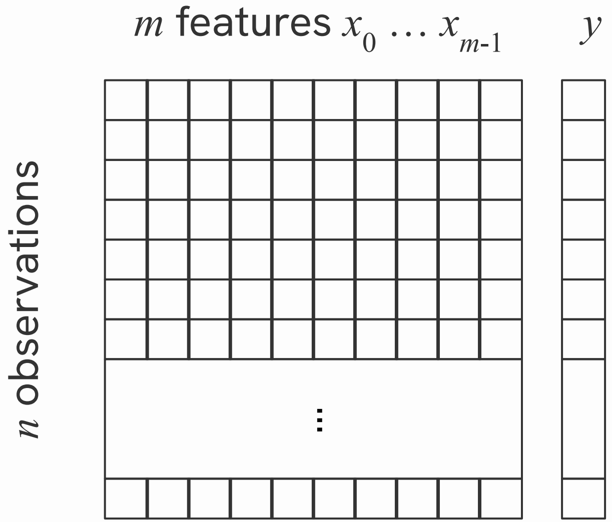
Impact Charts: Use a Collection of ML Techniques to Answer Social Science Questions

- Impact charts center explanation, not prediction.
- How does one feature variable impact a target variable *independent of other features*.

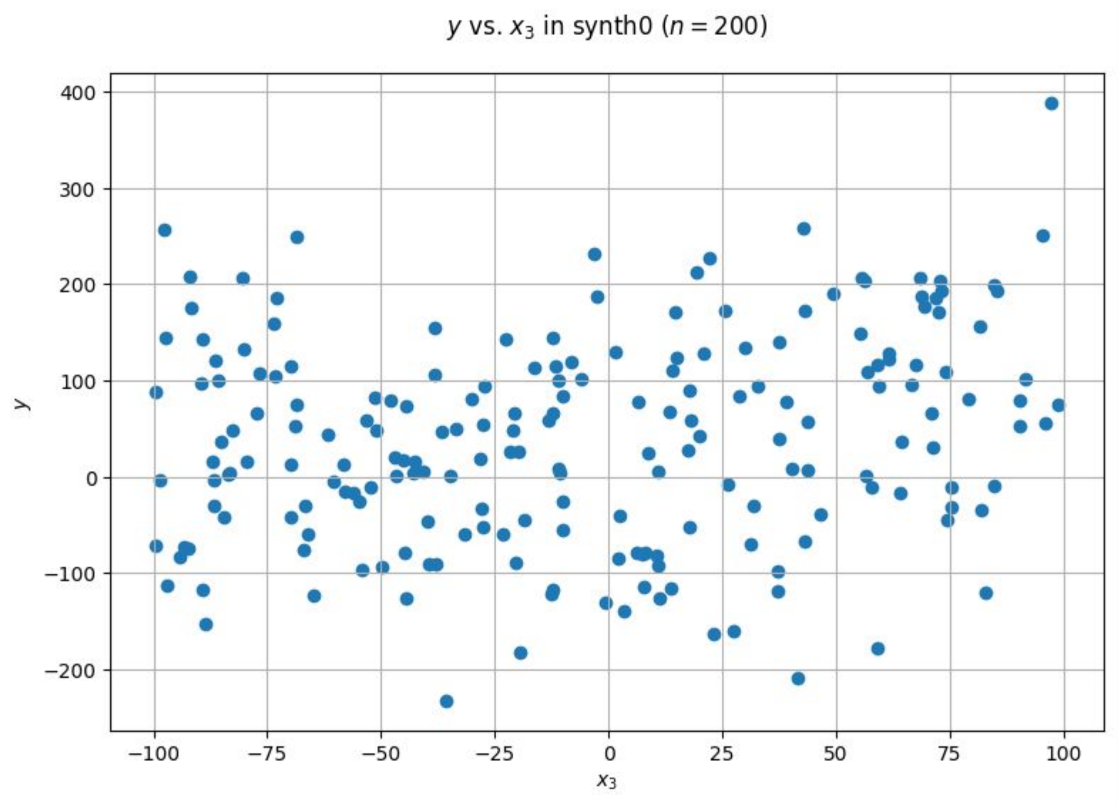
- Additional properties of impact charts:
  - Non-parametric: shapes are not known in advance
  - Allow many notions of error (e.g. MSE, MRE, ...)
  - Minimal assumptions, e.g. homoscedasticity not required.
  - Work well on aggregate data
    - Maintain more privacy at the individual level
    - Eliminate the need to impute individual features

The Problem

How is y Related to x_i?



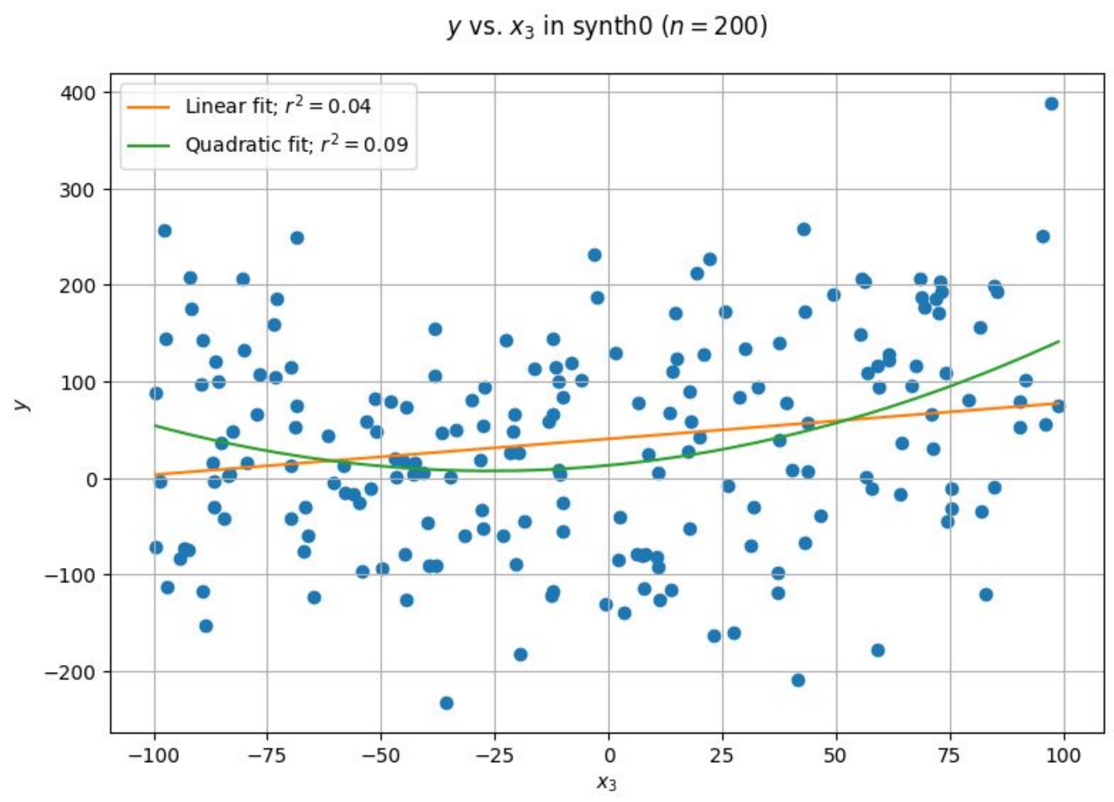
Exploratory Data Analysis: Scatter Plot



Why is the Scatter Plot Unsatisfying?

- There is no visually apparent pattern.
- But maybe y is impacted by x_3.
  - y might also be influenced by many other x_i.
  - We might not have observed all the relevant x_i.
  - There might be a lot of noise.
  - Some combination of the above.

Add Regression: Linear and Quadratic

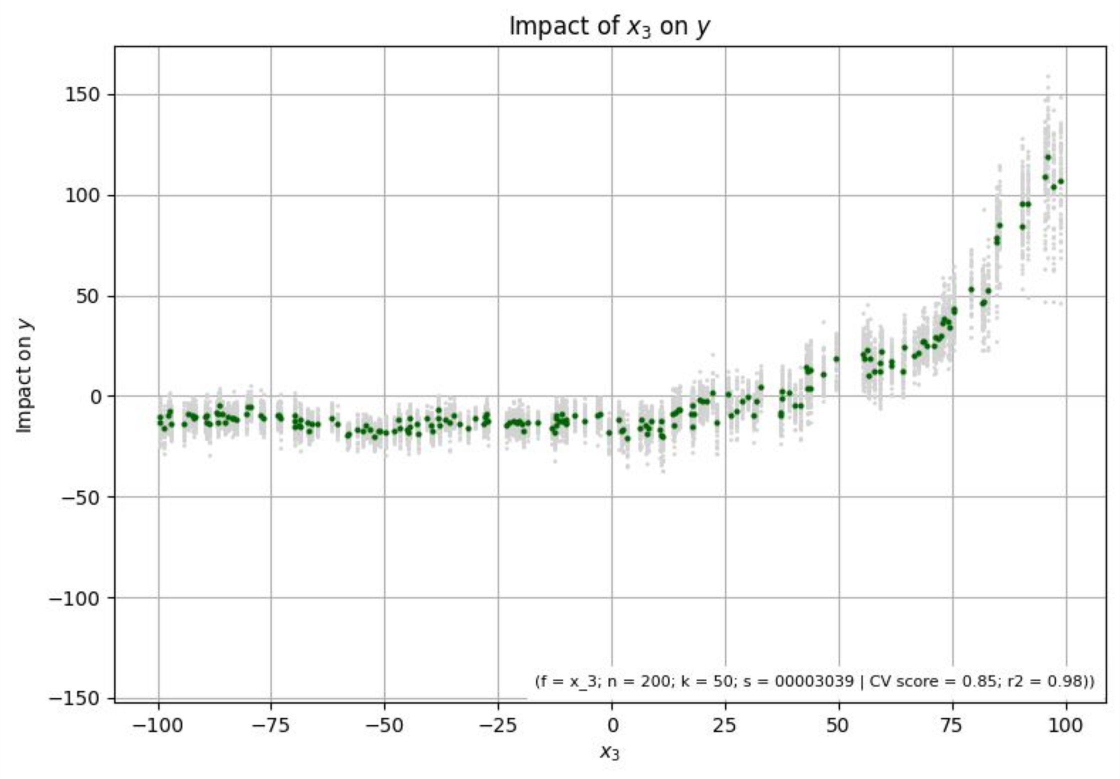


$y = ax_3 + b$

$y = ax_3^{\,2} + bx_3 + \mathrm{c}$

Still no real visual clues.

What if We Could See Impact Like This?



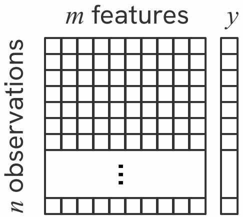
Methodology

How Do We Construct Impact Charts?

- Build an ensemble of k machine learning models
    - Bagging approach
    - Each model is on a sample of the original data
- Use Shapley values to compute the impact of each observation of each feature on the corresponding target
    - Impacts of features of an observation, by construction, sum to the difference between the training set mean and the model prediction for the observation
- Plot both the distribution and the mean impact vs. the observed value

One Model; Impact of One Observed Feature Value

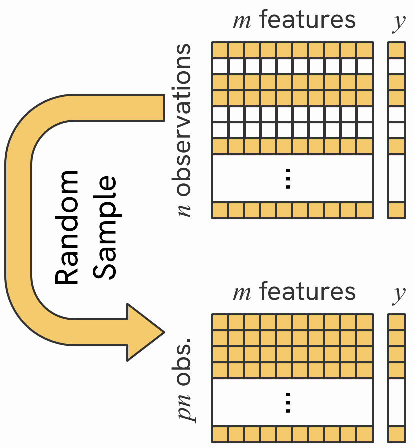
Given n observations of m features $X = (x_0, \ldots x_{m-1})$ and a target y.



One Model; Impact of One Observed Feature Value

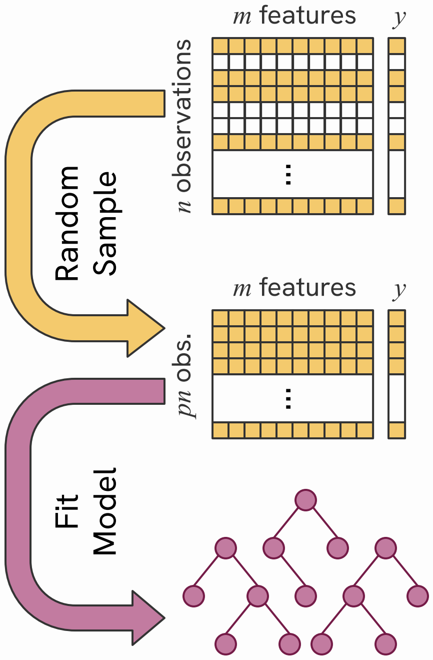
- Given n observations of m features $X = (x_0, \dots x_{m-1})$ and a target y.

- Take a random sample of the observations with e.g. $p = 80\%$.



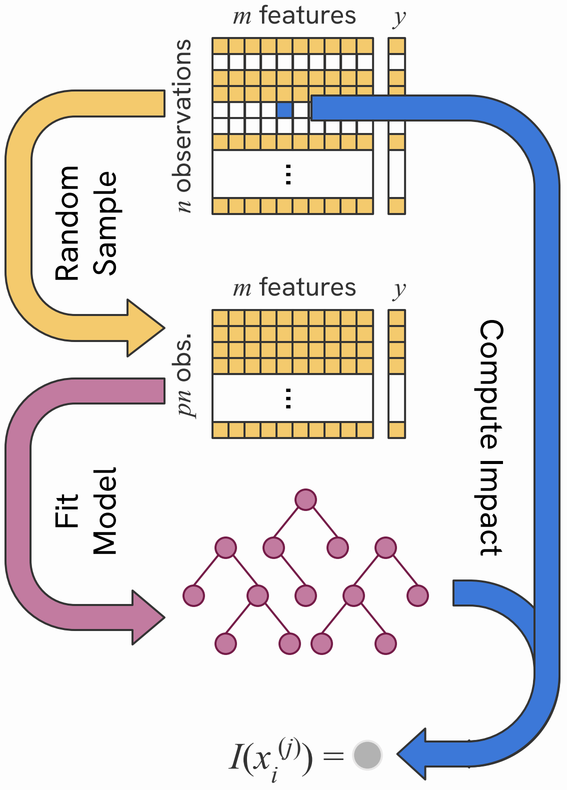
One Model; Impact of One Observed Feature Value

- Given n observations of m features $X = (x_0, \dots x_{m-1})$ and a target y.
- Take a random sample of the observations with *e.g.* $p = 80\%$.
- Fit a model on the random sample to predict y from X.



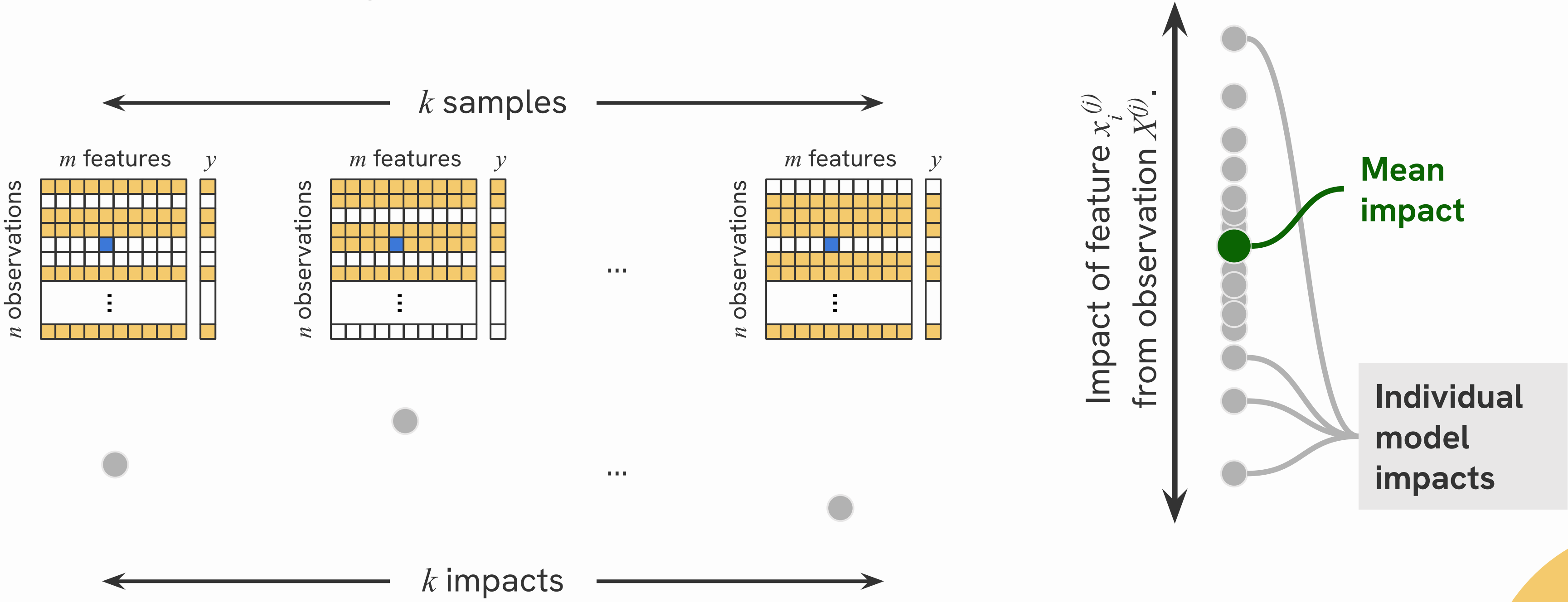
One Model; Impact of One Observed Feature Value

- Given n observations of m features $X = (x_0, \dots x_{m-1})$ and a target y.
- Take a random sample of the observations with *e.g.* $p = 80\%$.
- Fit a model on the random sample to predict y from X.
- Compute the impact $I(x_i^{(j)})$ of the j'th observation of the i'th feature (Shapley value)



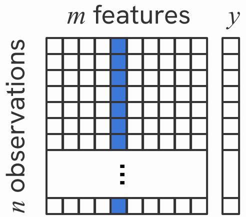
k Models; Impact of One Observed Feature Value

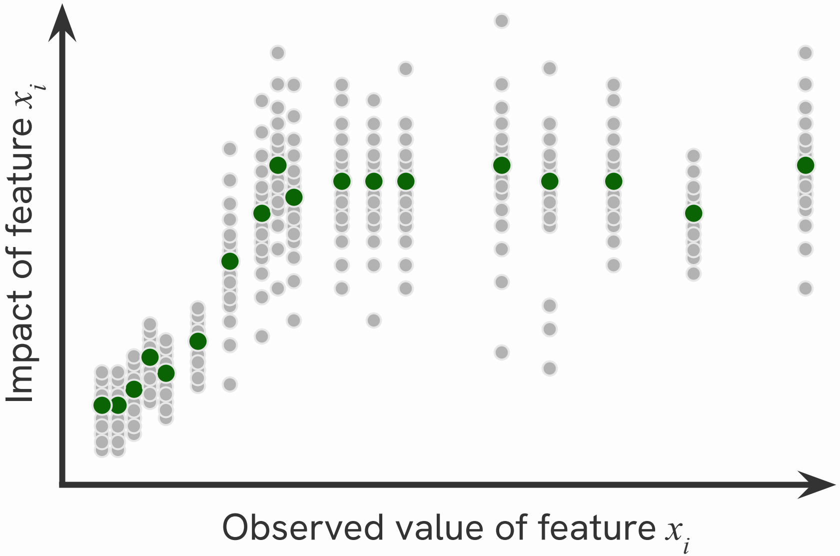
- Repeat the sampling, fitting, and impact calculation for k independent samples.
- Plot the distribution of k impacts vertically.
- Plot the mean impact.



Constructing an Impact Chart for Feature x_i

- Repeat prior multi-sample step for all *n* observations
- Plot impact distribution vertically vs. corresponding value of feature x_i
- Final chart shows how values of x_i impact y including a distribution
- Repeat to chart impact of other m - 1 features.





Validation on Synthetic Data

Synthetic Data: Impact is Known by Design

$$y = \sum_{i=0}^{4} t_i + N(0, 10)$$

y is sum of five terms, each of which is a function of one x_i, plus Gaussian noise.

$$t_0 = x_0;$$

linear in x_0.

$$t_1 = 100 \cdot \left(1 - 2\left(\frac{x_1}{100}\right)^2\right);$$

quadratic in x_1.

$$t_2 = 100 \cdot \sin\left(2\pi \frac{x_2}{100}\right);$$

sinusoidal in x_2.

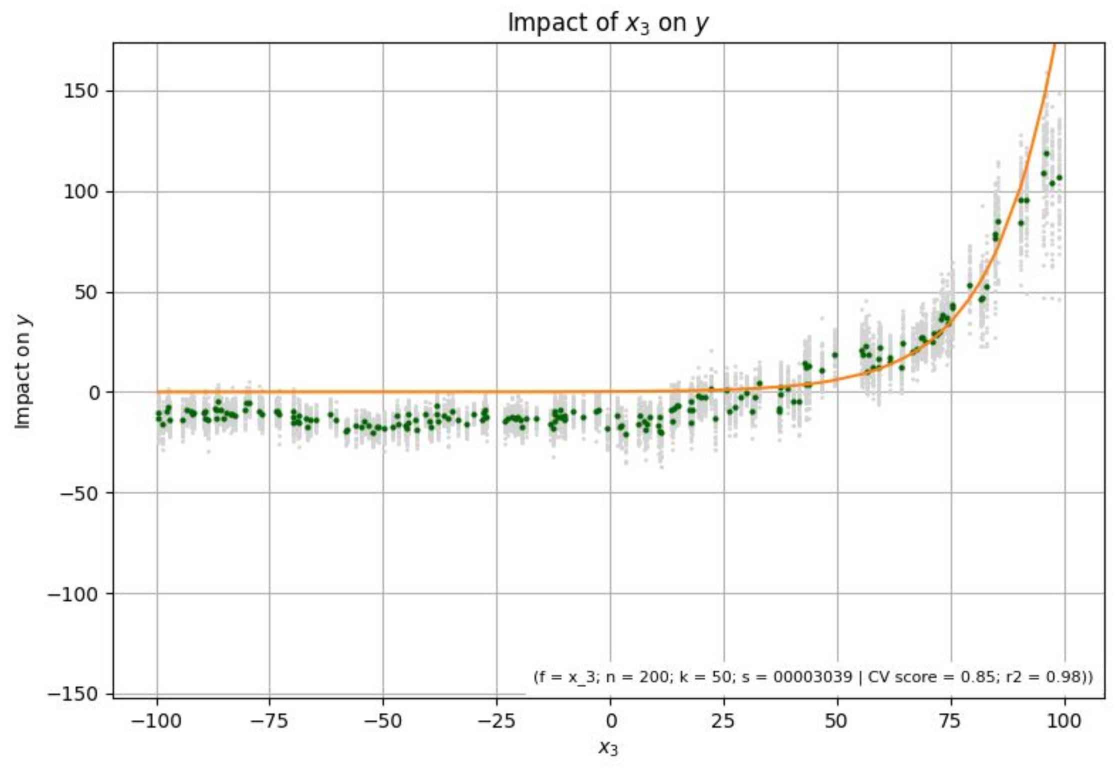
$$t_3 = 200 \cdot \exp\left[7\left(\frac{x_3}{100} - 1\right)\right];$$

exponential in x_3.

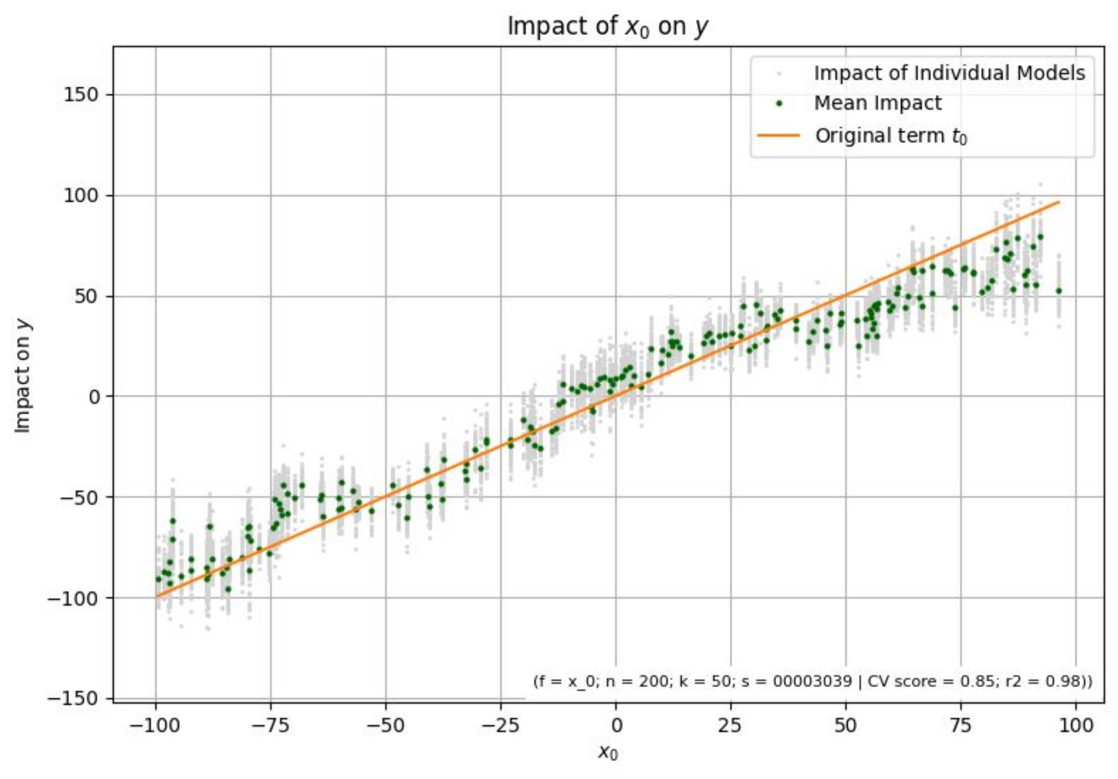
$$t_4 = 0.$$

unaffected by x_4.

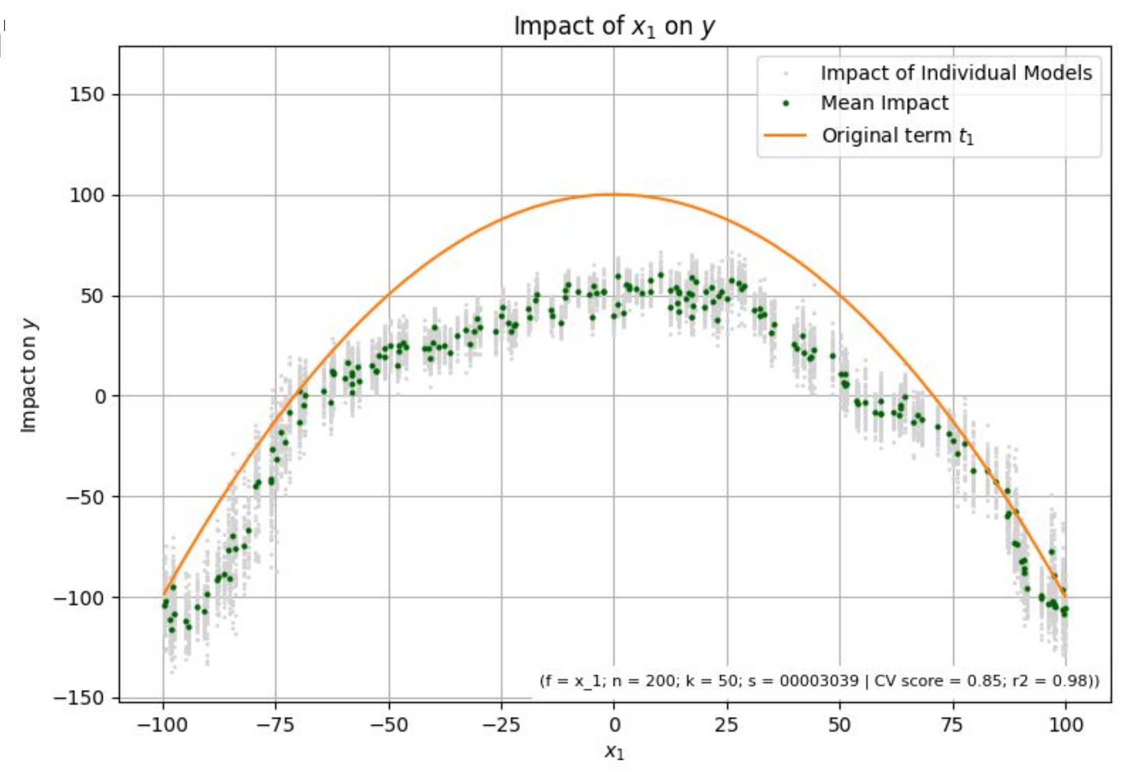
Impact Chart and Actual Impact of x_3



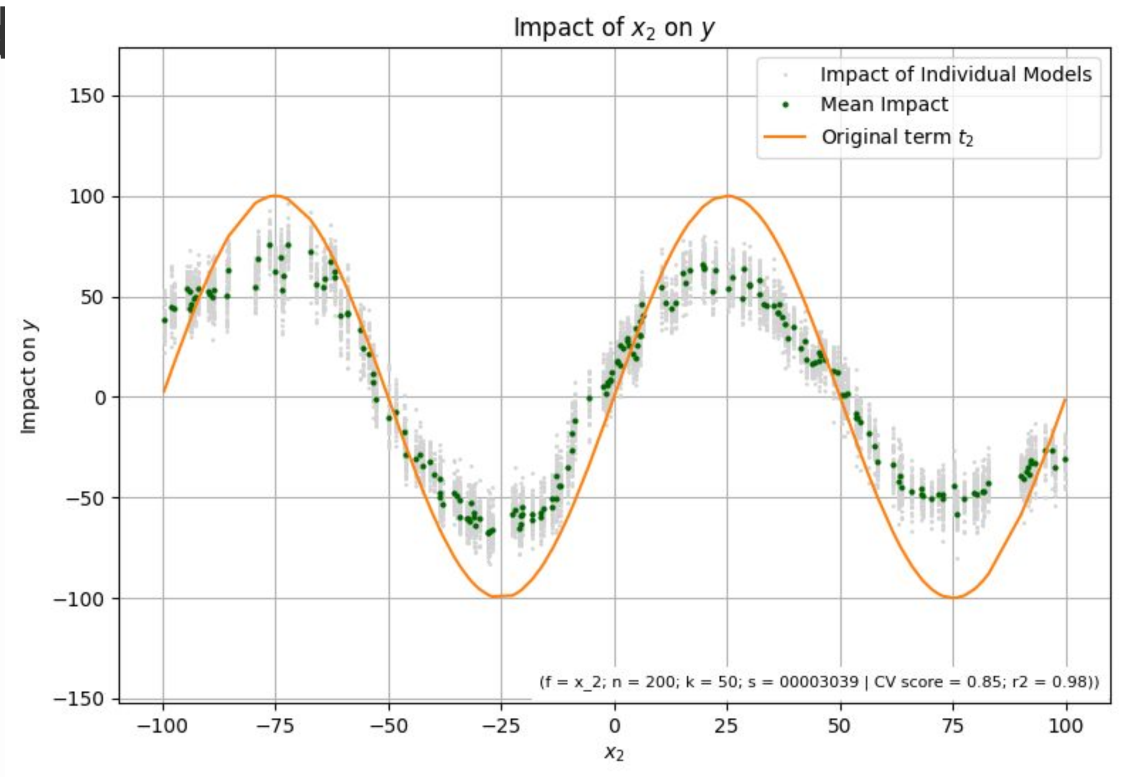
Impact Chart and Actual Impact of x_0 (Linear)



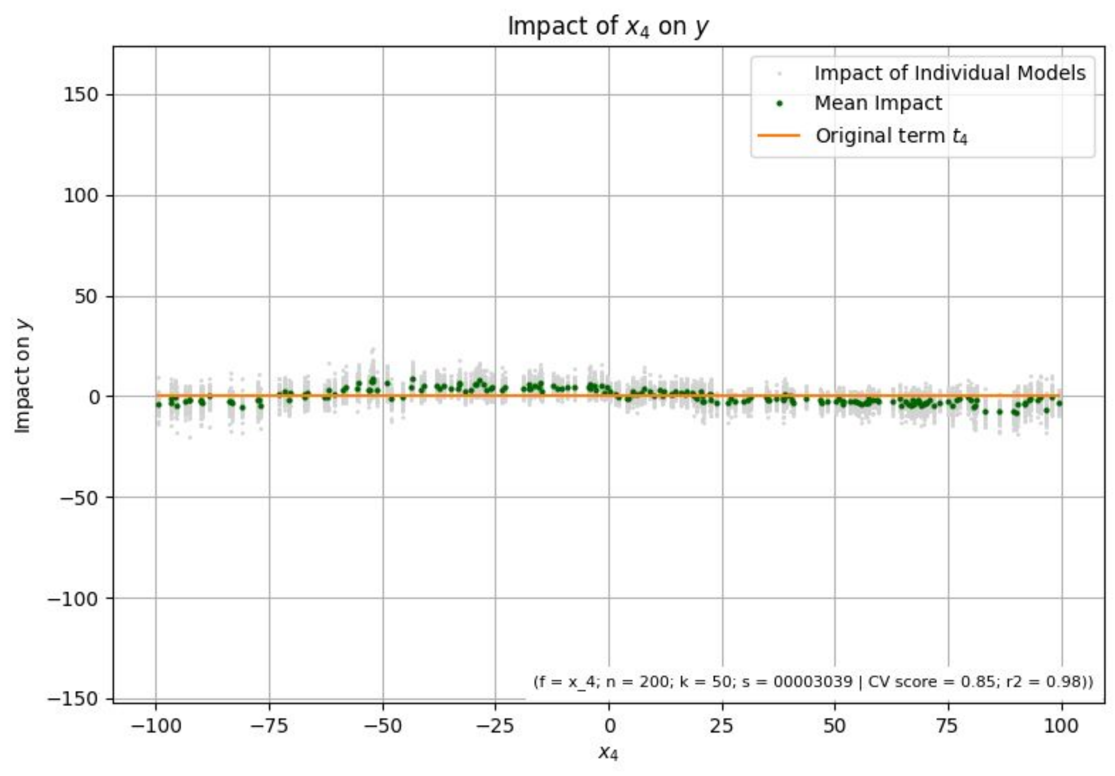
Impact Chart and Actual Impact of x_1 (Quadra



Impact Chart and Actual Impact of x_2 (Sinusoid



Impact Chart and Actual Impact of x_4 (None)



Impact Chart Code

Find the impactchart code in the usual places

- GitHub

  https://github.com/impactchart/impactchart

- PyPi

  ```
  pip install impactchart
  ```

Generate Impact Charts in 3-5 Lines of Python

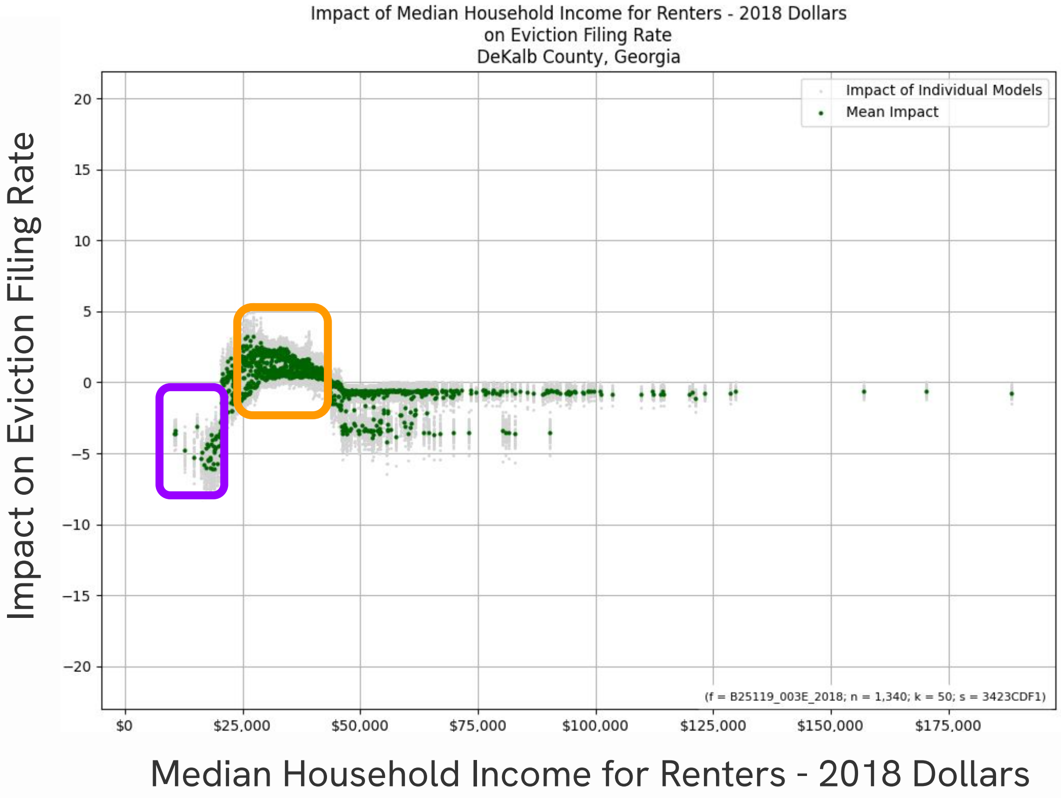
```python
from impactchart.model import XGBoostImpactModel

X, y = my_data()

# Construct and fit the impact chart model:
impact_model = XGBoostImpactModel()
impact_model.fit(X, y)

# Plot the charts. The return value is a dictionary
# with one chart per column of X.
impact_charts = impact_model.impact_charts(X)
```

A Real World Example

Example: Eviction Impact Charts

- The impact of income, race, and ethnicity on eviction rates in DeKalb County, GA, USA:
    - $n = 1{,}340$ census tract-level observations over 10 years (2009-2018)
    - Features are median income, percent of population identifying as [white | Black | Asian | Hispanic or Latino, ...]. (U.S. Census American Community Survey data from 2009-2018.)
    - Target is eviction rate in eviction filings per 100 renters per year (Princeton University Eviction Lab data)
    - $k = 50$ models on 80% samples using XGBoost with optimized hyperparameters

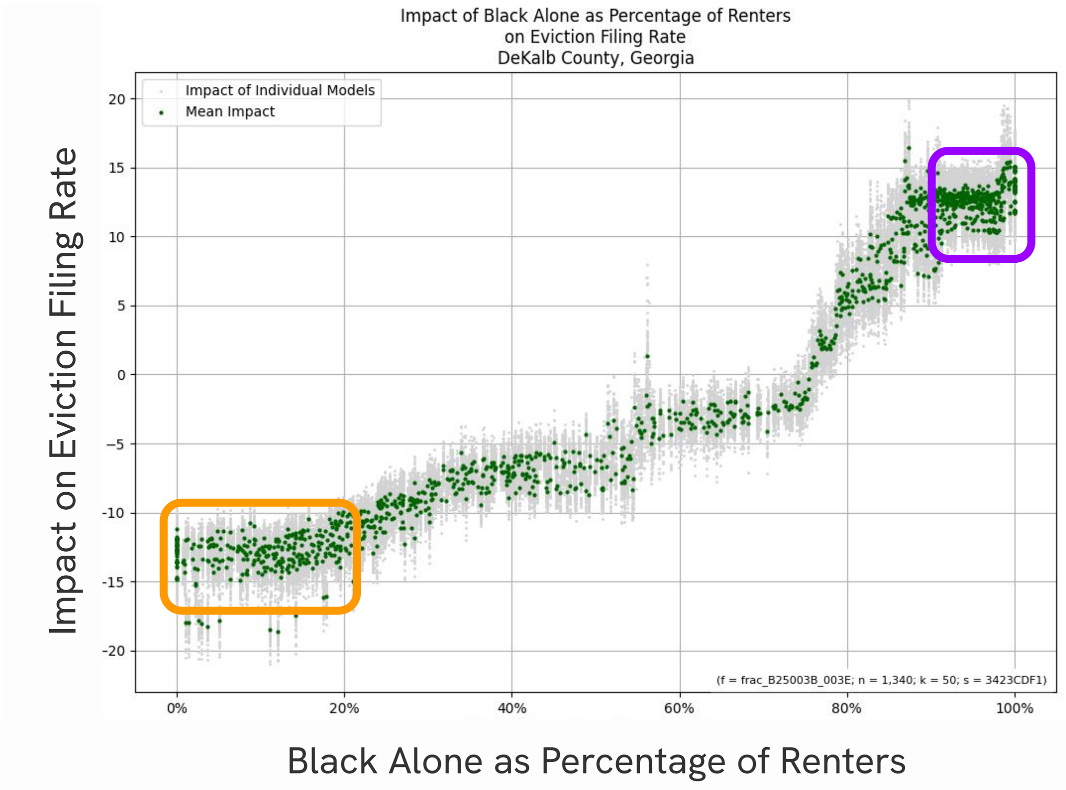
Example: Impact of Income on Eviction



- Hypothesis: low income leads to higher eviction.
- Impact chart results: reality is more complex.

- Census tracts where median renters are among **the working poor** have eviction rates up to 3 points higher than otherwise similar tracts.
- Census tracts where median renter's income is **the absolute lowest** have eviction rates up to 6 points lower than otherwise similar tracts.

Example: Impact of Blackness on Eviction



- Census tracts that have **low Black population (<20%)** have eviction rates 10-15 points lower than relative to otherwise similar tracts.
- Census tracts that have **high Black population (>90%)** have eviction rates 10-15 points higher.
- Overall range of 30 points difference vs. < 10 points for income.
- Recall: impacts are independent and additive.
- Compare to scatter plots/regression. See paper.

Paper with More Details And Examples

Darren Erik Vengroff. 2024. Impact Charts: A Tool for Identifying Systematic Bias in Social Systems and Data. In *FAccT '24, June 2024*.

https://facctconference.org/static/papers24/facct24-80.pdf

Future Work

- Causal models and impact chart resolution
- Interactive impact charts
- Additional application areas

Thank You

https://github.com/impactchart/impactchart