Predicting Poverty from Satellite Imagery

Zotti LLC

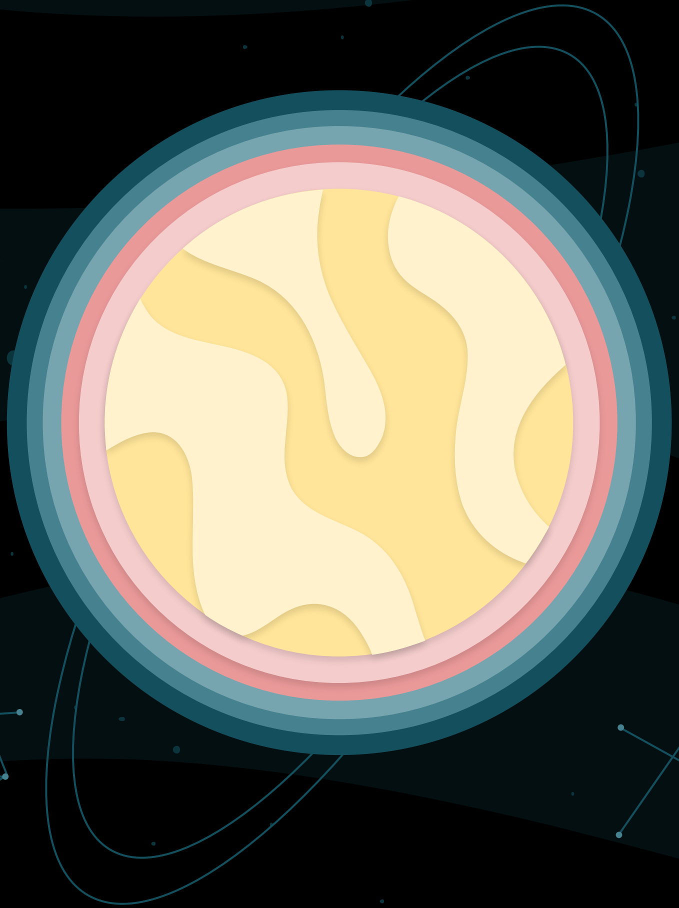
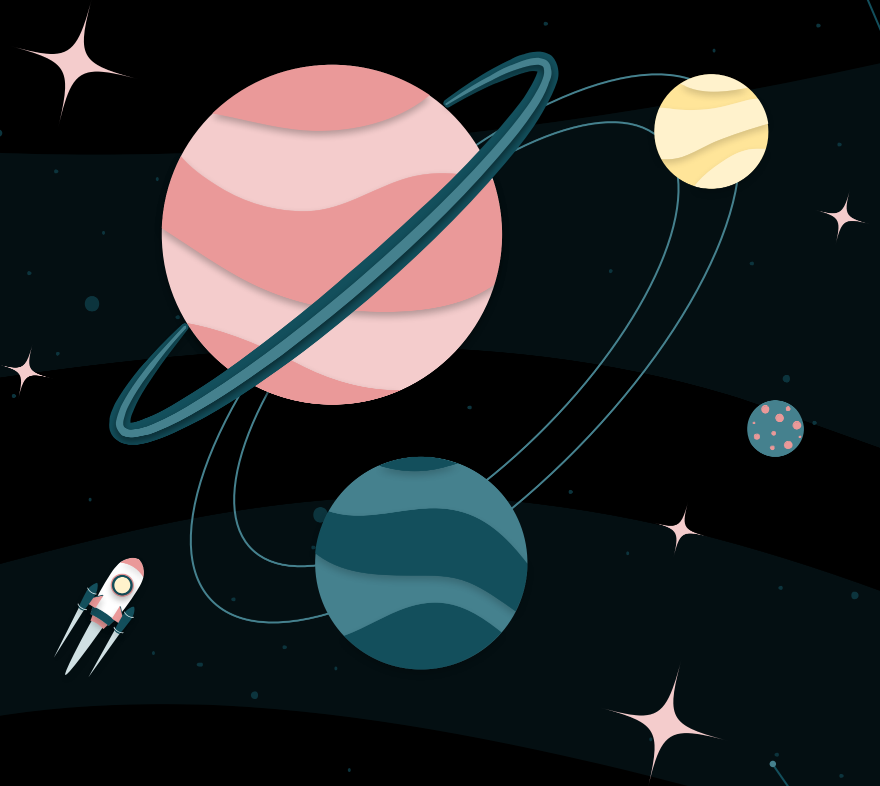


TABLE OF CONTENTS

01

Problem Statement



Can satellite imagery be used to predict poverty in US cities?

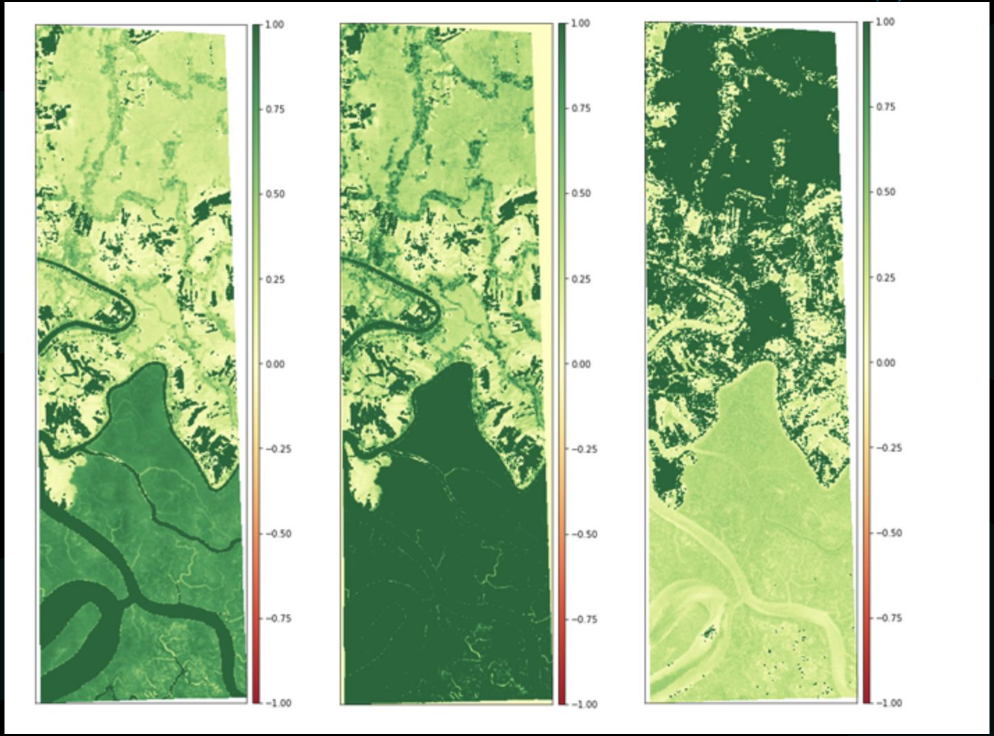
We will look at DC and Chicago satellite imagery to find out...

Why?

- Census only once every 10 years
- Hard to capture people without permanent addresses
- Does poverty look different in different cities?

Background on satellite imagery: Nightlights

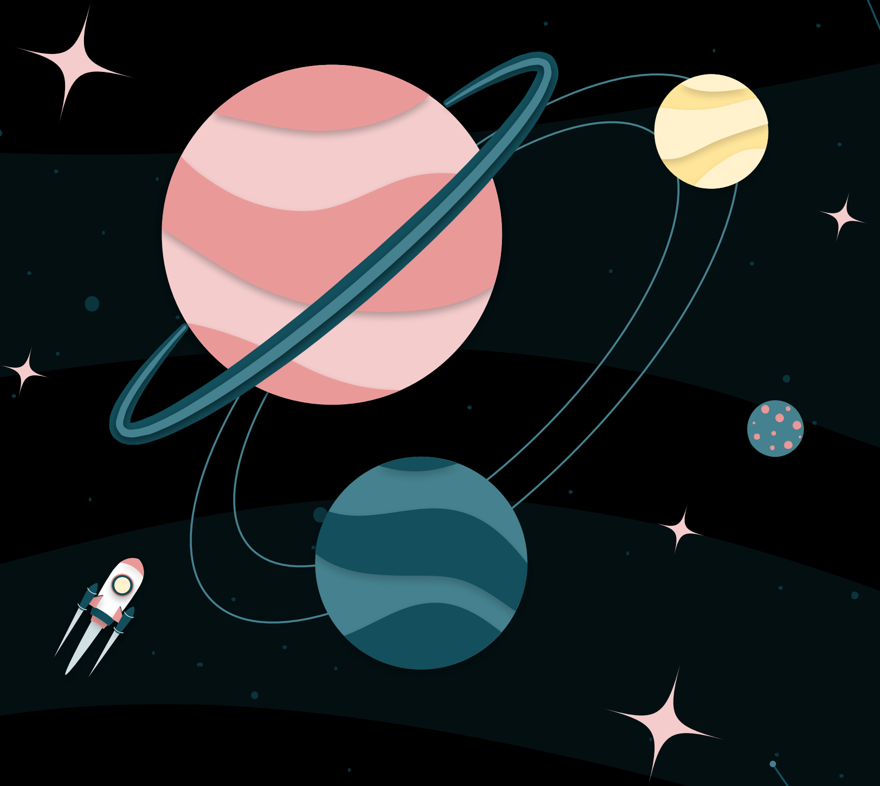
Background on satellite imagery: Agriculture



Three different ways of calculating sunlight reflected by plants, from this blog

02

Acquiring the Data



Sources

Google Earth

Used to acquire both DC and Chicago imagery

Censusapi

Income data for each census block group

Data Portals

Each city had data portals to download shapefiles

New Libraries

Geopandas

Allows coders to work with shapefiles and polygons

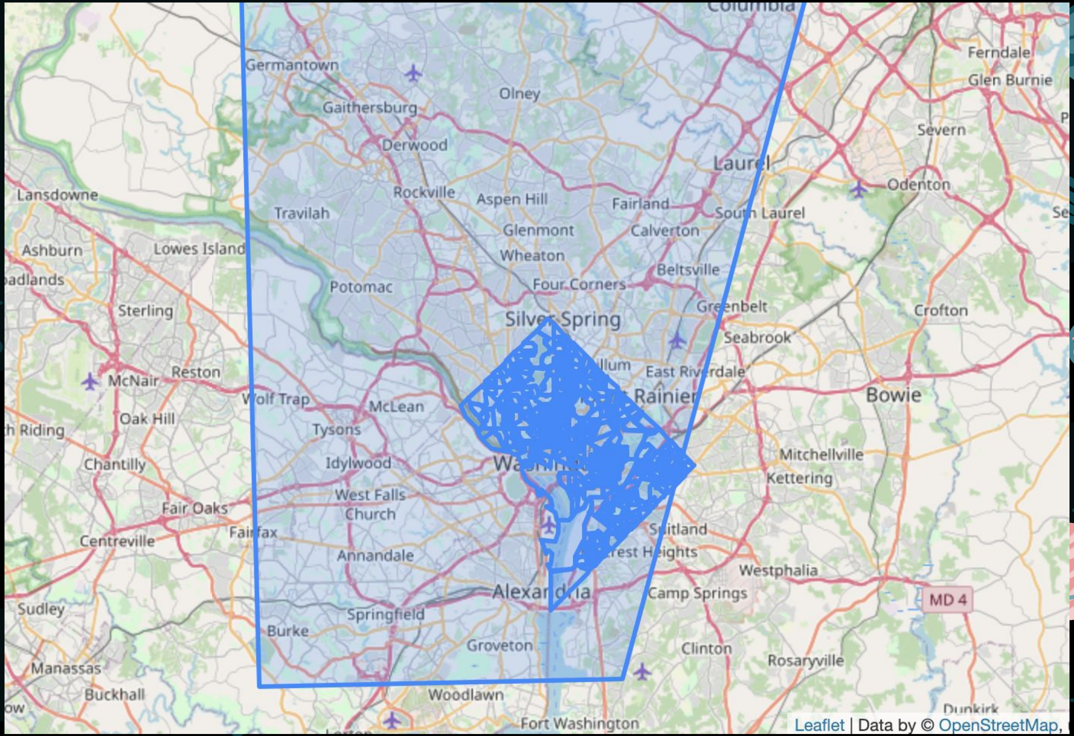
Rasterio

Allows coders to work with imagery - like pictures but with many more spectral bands and geographic metadata

Sentinel 2

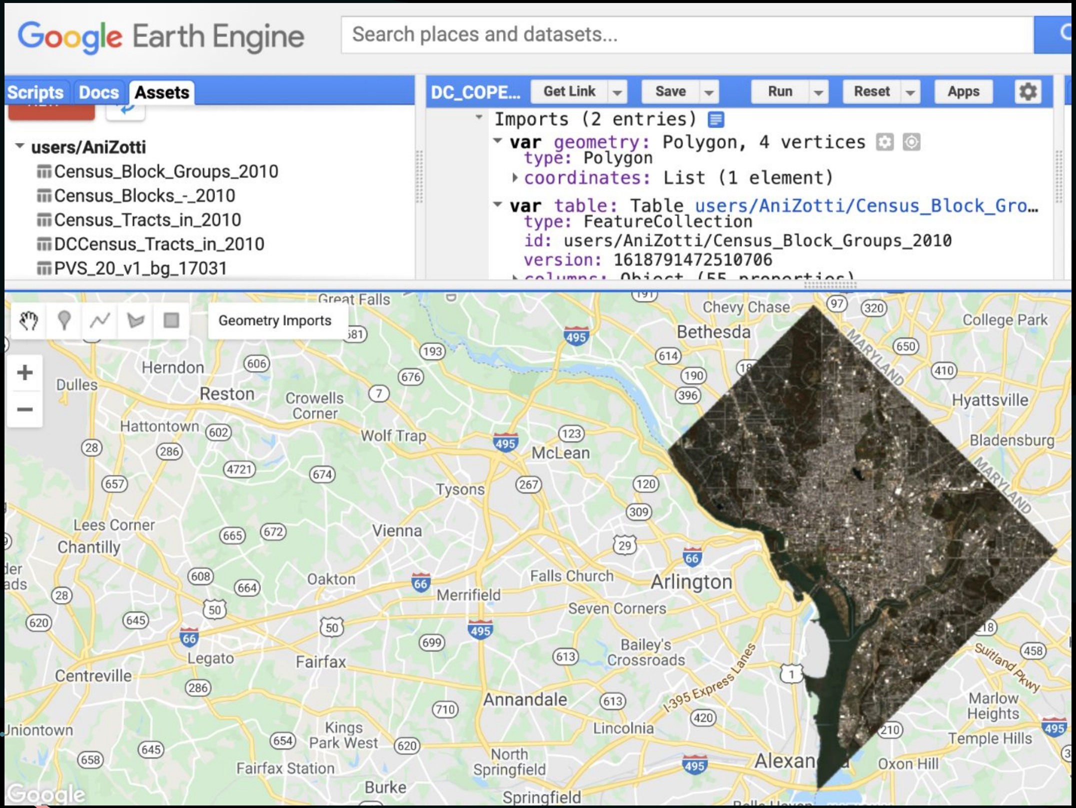
- European Space Agency
- Actually 2 different satellites
- Circles the Earth once every 5 days between the two
- 10m resolution
- 13 spectral bands

Attempt 1: regularly plotting GIS data is important!



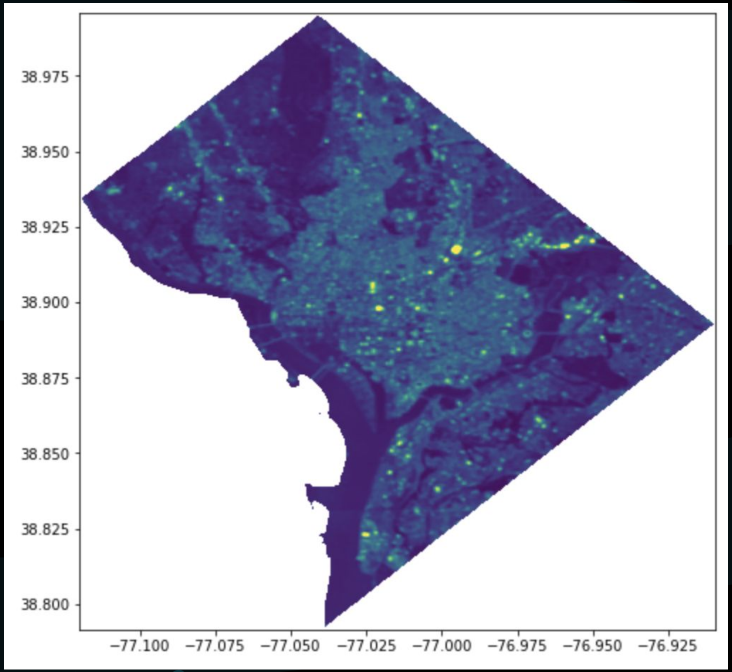
Using SentinelAPI

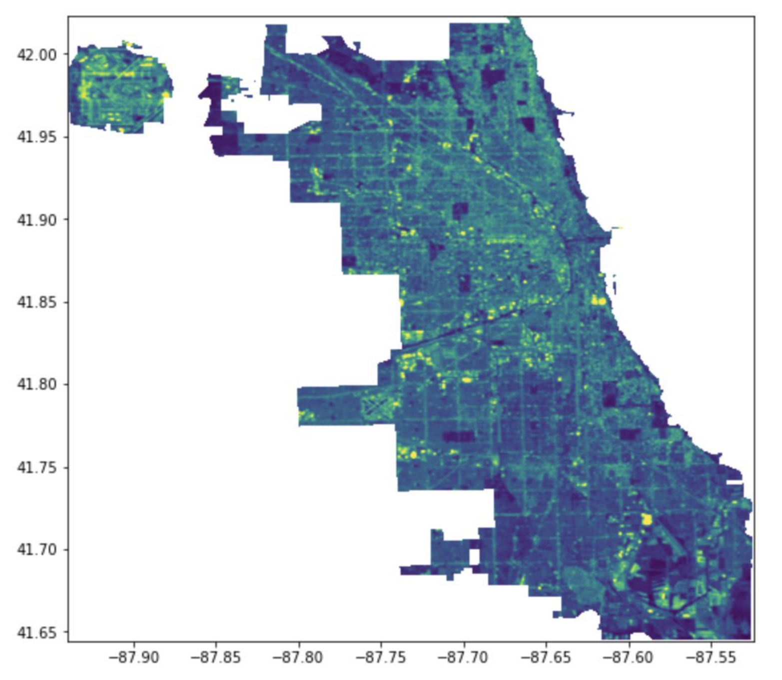
Google Earth Engine



- Building a cloud mask
- Taking the median value for each pixel across each spectral band over a specific timeframe
- Clipping the image by a certain geometric boundary
- Exporting the image to Google Drive

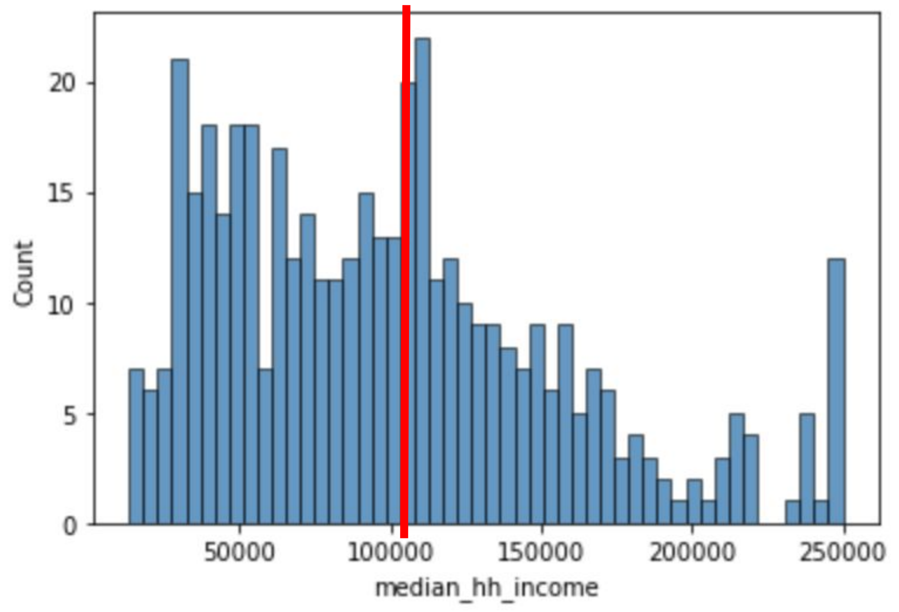
Imagery in Jupyter



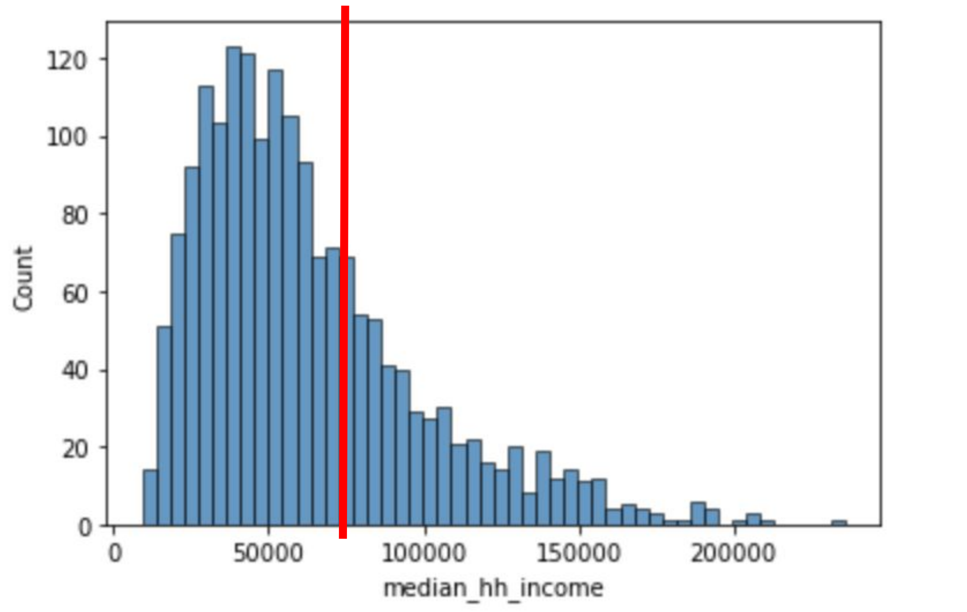


Income Distribution

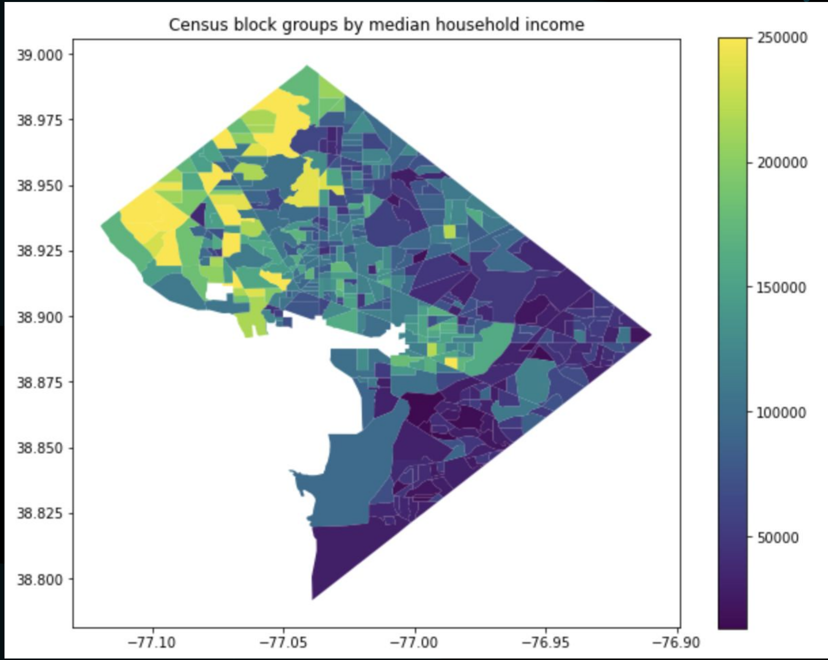
DC

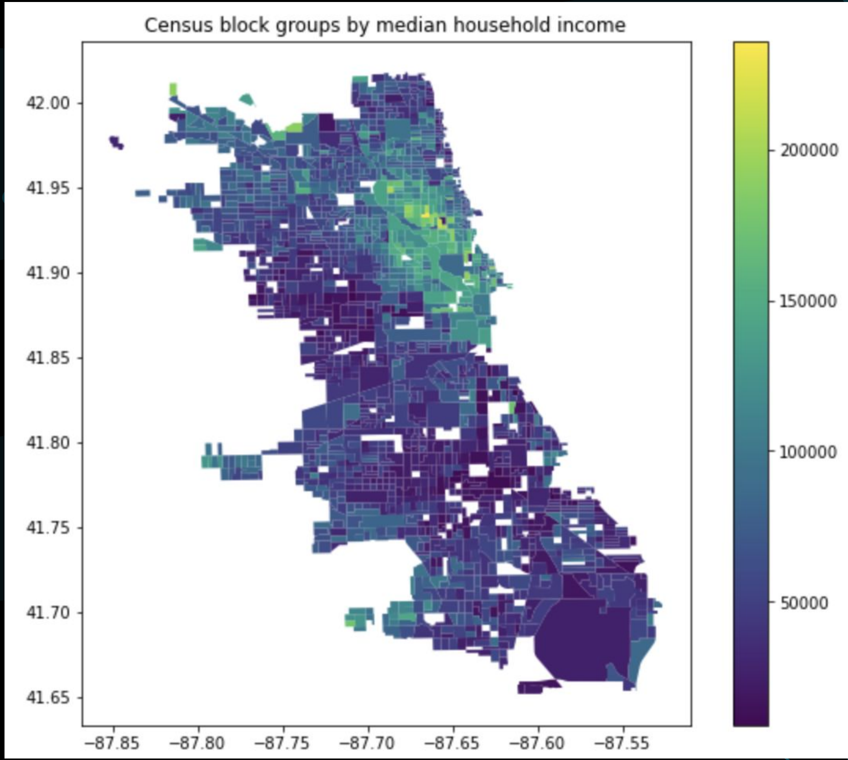


Chicago



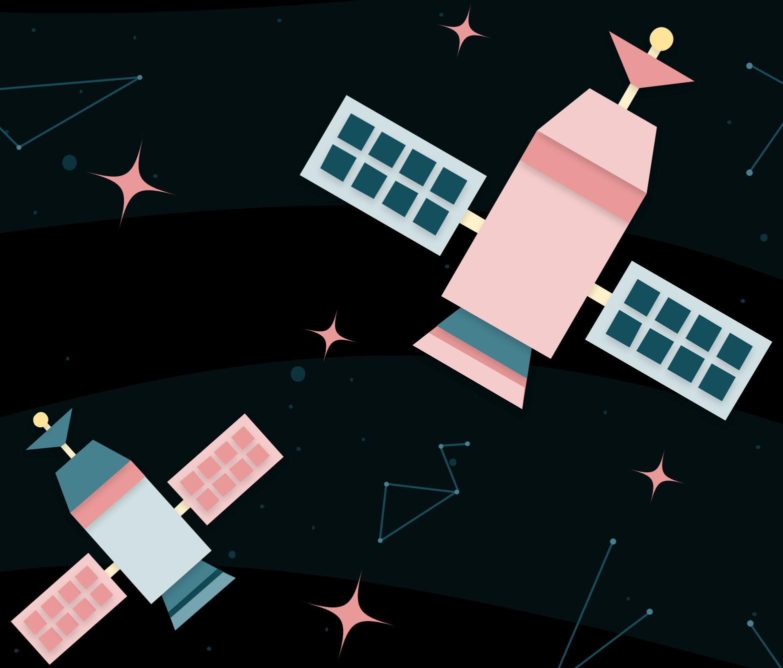
Imagery in Jupyter





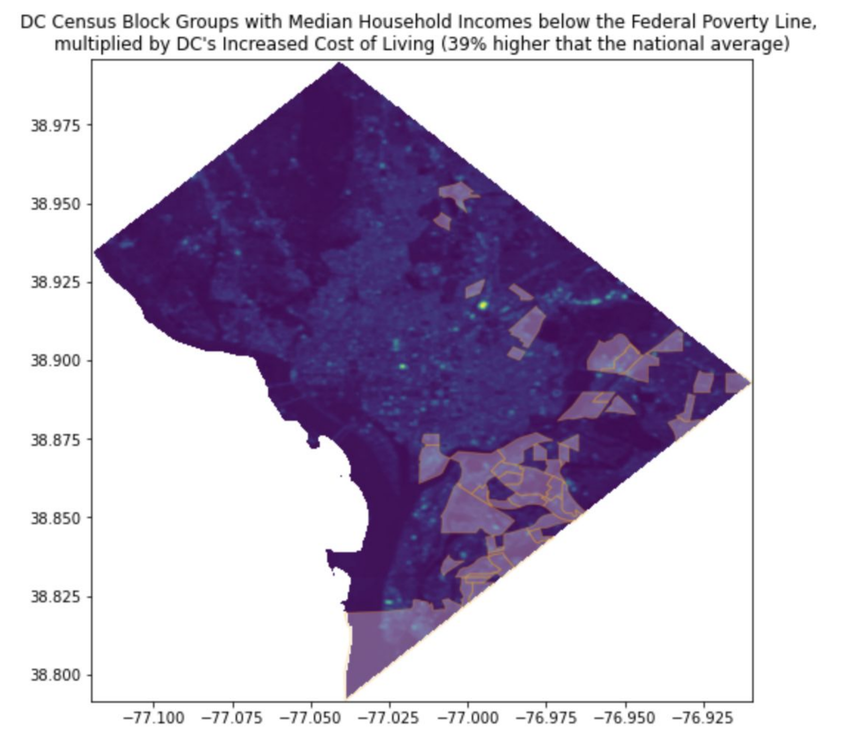
Poverty

- Federal Poverty Level for a Family of Four: 26,500
- DC's Cost of Living over the National Average: 39%
- Chicago's Cost of Living over the National Average: 23%



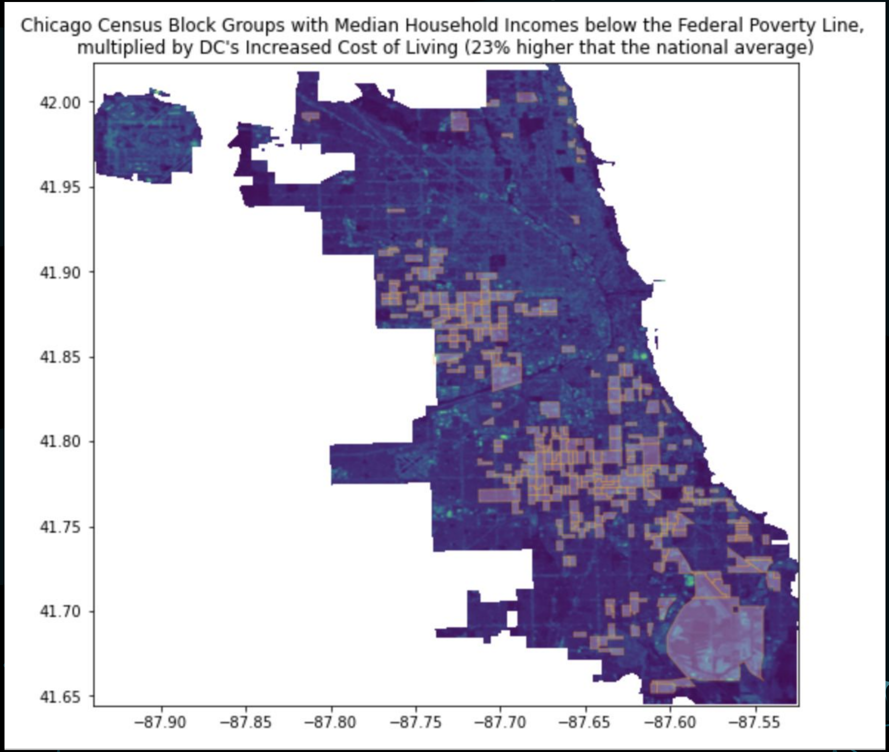
Pulling everything together

DC



Baseline: 94%

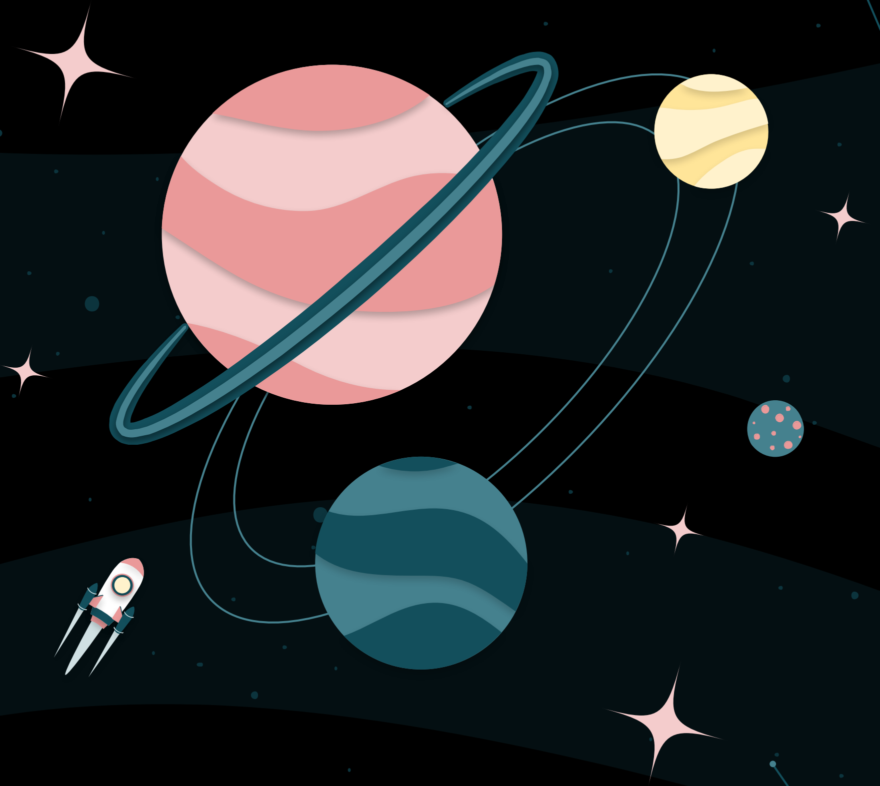
Chicago



Baseline: 93%

03

Modeling



Preprocessing: “Rasterizing” the polygons

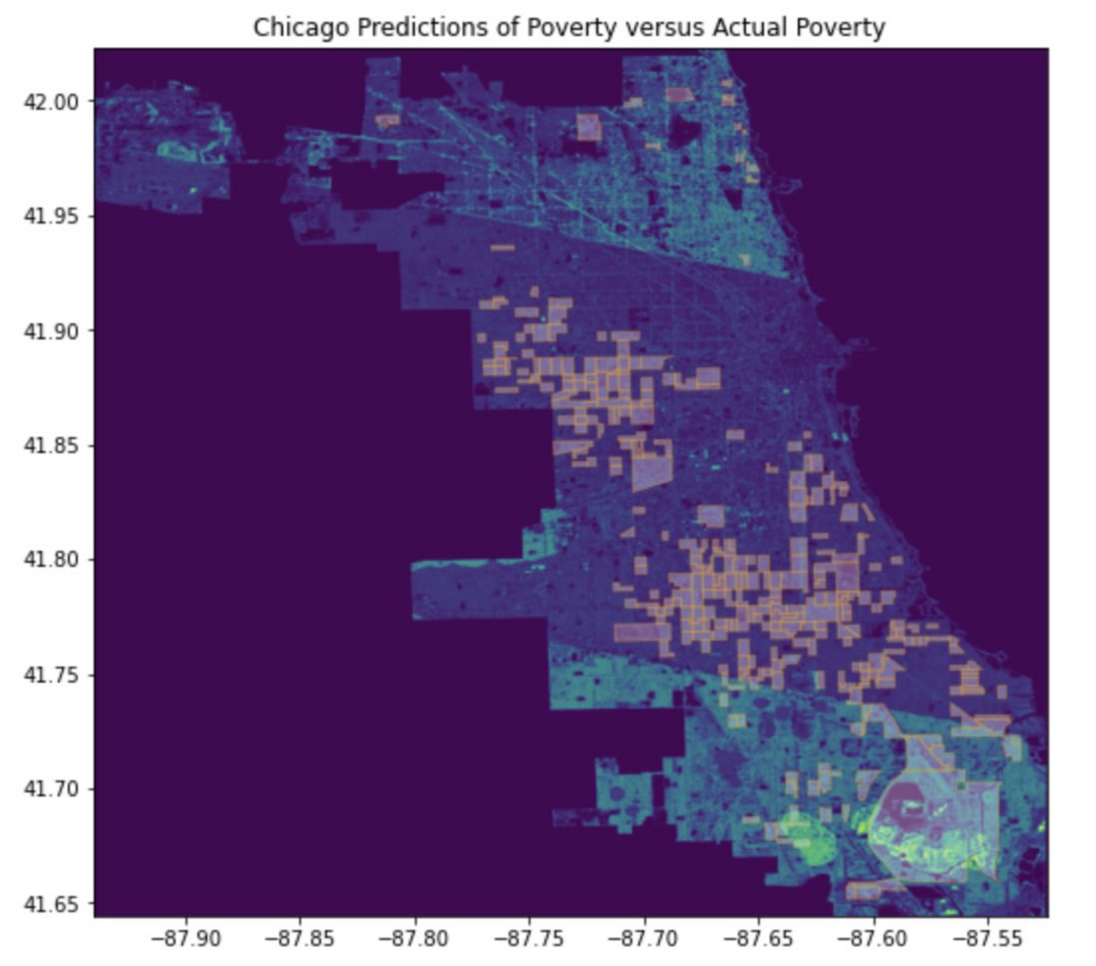
DC

Chicago

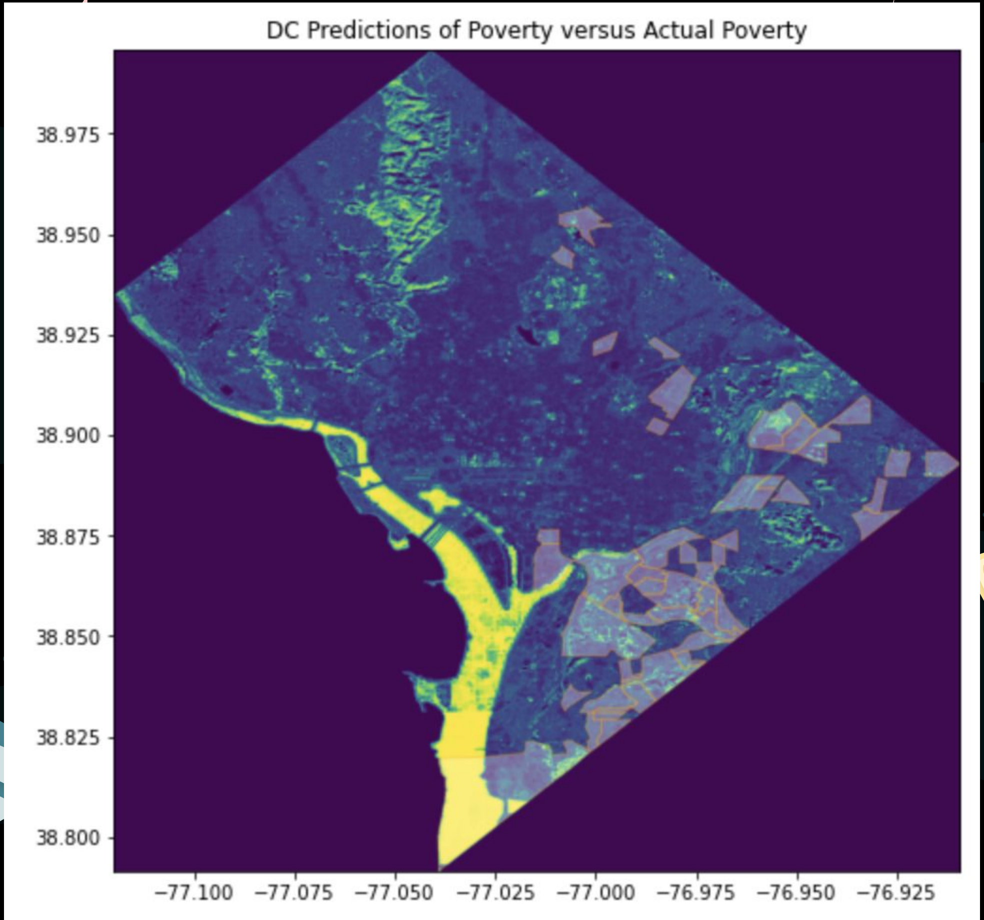
The models

Models	Baseline Accuracy*	Precision*	Recall*	Accuracy*	New city?
Training on DC data, predicting Chicago poverty	**0.943**	**0.68**	**0.31**	**0.953**	✓
Training on DC data with artificially balanced classes	**0.943**	**0.14**	**0.16**	**0.89**	✗
Training on Chicago data, predicting DC poverty	**0.930**	**0.77**	**0.057**	**0.9348**	✓
Training on Chicago data with artificially balanced classes	**0.930**	**0.34**	**0.58**	**0.893**	✗

Results

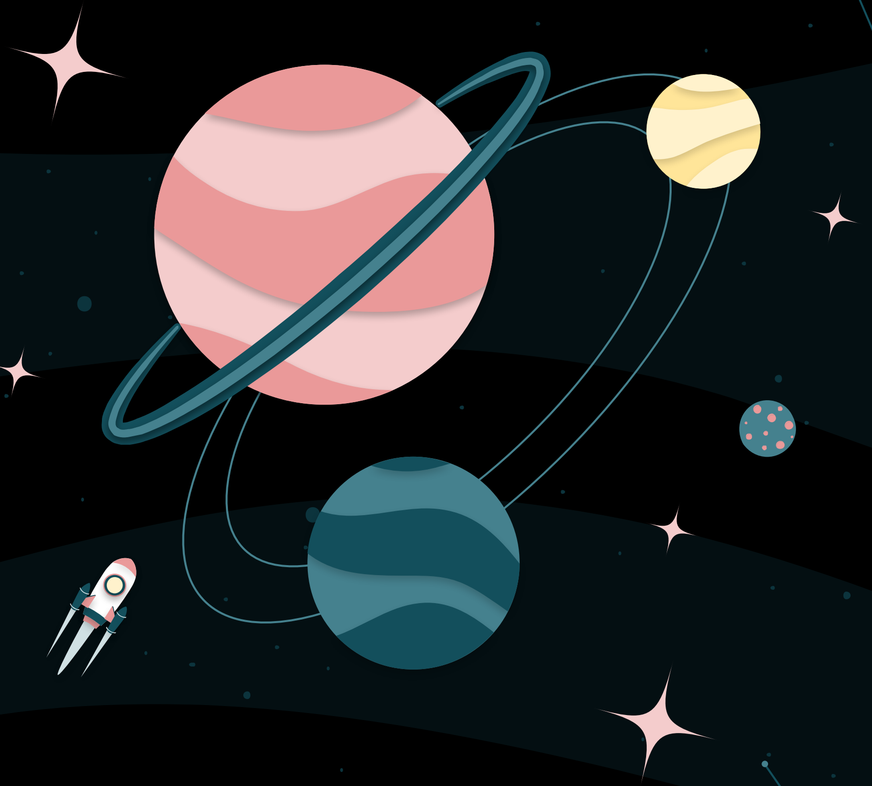

Chicago Predictions of Poverty versus Actual Poverty
42.00
41.95
41.90
41.85
41.80
41.75
41.70
41.65
−87.90
−87.85
−87.80
−87.75
−87.70
−87.65
−87.60
−87.55

Other Results



04

Conclusions



Recommendations

More data

Thousands of satellite images

More labels

Open Data Portals have shapefiles for buildings, parks, etc

More compute

Required if using more raster files

Takeaways

Workflow

A success!

Accuracy

Needs work

Different cities

Potentially!