

## **Data Analytics Final Project**

**Topic:** Analyzing Float Glass Plants by Electrical Rates

**Contributors:** Faris Alquaddoomi, Ethan Frisch, Megan Morrissey, Ajay Parthasarathy

### **Problem Statement**

Glass factory float lines generate significant amounts of wasted heat, which represents an opportunity for heat reclamation. The problem lies in converting the heat to useful electricity with an economically viable waste heat recovery system. Glass factories currently source electricity from the grid at competitive rates for industrial consumption. For our solution to be cost-competitive with electricity generated from fossil fuel plants, it must supply electricity at a lower comparative price.

We intend to explore methods by which we can supply electricity at a lower comparative price to fossil fuel plants. The first method is based on the price that we can generate electricity with our thermophotovoltaic (TPV) waste heat recovery system. Some estimates show electricity rates for TPVs at around \$0.06/kwh - which is already half the price of fossil fuel-generated electricity (quoted at an average of \$0.12/kwh). However, TPVs are expensive to make and the price of the system would factor into the price of electricity produced by the system. We intend to explore the cost-competitiveness of our system using datasets on state-level electrical prices compared to our best-case-scenario for TPV electricity rates (\$0.06/kwh).

### **How will the problem be tackled and solved?**

We will generate maps of electrical price, float line location, and float line number at each glass site. These maps will rely on geopandas and geoplots for visualization. We will perform K-means cluster analysis on the maps to identify trends in glass factory locations. We will use visualization methods like a histogram of electrical price for float glass plants to identify what we can charge in each geographical area. The primary assumption is that float glass plants are paying the state-wide industrial rate for electricity and not more or less.

### **What are the parameters around your problem statement to make it simple enough to solve but not trivial?**

The first parameter around our problem statement is cleaning and filtering out-of-scope glass plants from our datasets to constrain our analysis to US-based plants. The next parameter is ensuring the dataset value types are correct for the purpose of processing in our model - e.g. each column has a list of properly formatted strings or floats for the purpose of geolocation. The final parameter is actually performing the geolocation and cluster analysis on our datasets and creating clear visualizations from the results. These parameters are complex but achievable and require using various libraries to simplify the solution. The solution is non-trivial in that it provides us with new insights into electricity price geographics and how they associate with float glass plant location.

## **Data Analytics Final Project**

**Topic:** Analyzing Float Glass Plants by Electrical Rates

**Contributors:** Faris Alquaddoomi, Ethan Frisch, Megan Morrissey, Ajay Parthasarathy

**What is the core business or research problem you are solving? Why is it important? This should be concise and clear.**

The core business problem we are trying to solve is to *determine how much we can charge for electricity at each float glass line we install our waste heat recovery system in.*

Because our waste heat recovery system generates electricity for the glass plants, we want to charge them for electricity at a rate lower than what they currently pay. In order to know how much we can charge to remain competitive with on-grid electrical prices, we need to find out how much is currently being paid at each float glass plant; this depends on the geographic location of each plant, since electricity price varies by state.

### **Dataset**

**Is the data collected and loaded appropriately**

Data collection is from official World of Glass datasets. Loading is done through the csv module in python. Datasets will be converted to csv files and read into the working environment as matrices. The original matrices will be cleaned and filtered for addition into pandas dataframes and subsequent addition into geopandas geo-dataframes.

**Does dataset have the capacity to address the question posed in the problem Statement**

The datasets include all major US-based float glass lines and the city they are located in plus industrial electricity prices for each state in 2019. These datasets are sufficient to answer the problem posed.

### **Data Cleaning**

**Was data ingested, sufficiently cleaned, in a format that makes it amenable for visualization, model building and analysis**

Datasets were properly ingested, cleaned, loaded, and formatted for the purpose of visualization, model building, and analysis. Specifically, the data is loaded below as csv files and processed into dataframes and eventually geo-dataframes. Irrelevant classes are discarded for the purpose of simplifying the datasets. New index values are contributed to the dataframes to supplement the analysis (e.g. adding “ ,US” to each city to constrict adding lat-long coordinates for only US-based cities).

## **Data Analytics Final Project**

**Topic:** Analyzing Float Glass Plants by Electrical Rates

**Contributors:** Faris Alquaddoomi, Ethan Frisch, Megan Morrissey, Ajay Parthasarathy

### **How were missing data handled**

Missing data did not exist between relevant rows and so we did not need to perform any interpolation or extrapolation of values. There was empty data past a certain row in our World of Glass dataset, but this was due to the original csv format. The empty data was simply dropped from the dataframe.

### **How was the data appropriately normalized or scaled**

The data for electrical prices was placed into 5 bins corresponding to 5 electrical price ranges.

### **How were the data initially collected? Units? Any metadata should be linked or mentioned.**

Data was collected from three sources: World of Glass dataset on glass plant information globally; Electrical Price rates from the US Electricity Information Administration; lat-lon coordinates from the Nominatim OpenStreetMap web database were pulled using reverse geocoding functions from the geopandas library.

Datasets are as follows:

- Electrical Prices: [https://www.eia.gov/electricity/monthly/xls/table\\_5\\_06\\_a.xlsx](https://www.eia.gov/electricity/monthly/xls/table_5_06_a.xlsx)
- Glass Plant Locations: <https://members.glass.org/cvweb/cgi-bin/msascartdll.dll/ProductInfo?productcd=WOGFLOAT>
- Lat-Lon Coordinates: <https://nominatim.openstreetmap.org/ui/search.html>

## **Data Visualization**

### **Create clear, labeled and appropriate visualizations**

We produced 3 primary types of maps: Choropleth base maps, Pointplot overplots on polyplots, and Pointplot overplots on the choropleth base maps. Each base map is given an appropriate legend and labeled according to the information it displays.

We also constructed a histogram for glass plant electricity prices in the US: their frequency and industrial electricity rate, assuming statewide industrial electricity prices are charged to the plant.

### **Visualizations should be made for exploratory phase of your analysis as well as for the insights and results from your model**

We created prototypes of the maps before overplotting. We also checked the pointplot maps before performing K-means clustering analysis. Overplotting was done once the prototypes were appropriately passed to the analysis.

## **Data Analytics Final Project**

**Topic:** Analyzing Float Glass Plants by Electrical Rates

**Contributors:** Faris Alquaddoomi, Ethan Frisch, Megan Morrissey, Ajay Parthasarathy

### **Model Building**

**What model did you use for making predictions or solving the problem**

We used a geo-dataframe with a geometry column for passing to the K-means cluster analysis function. K-means requires coordinates under the geometry class to predict the clustering. It also requires a predetermined number of clusters to search for; we determined the appropriate number of clusters using the Elbow Method. The Elbow Method is a visual method for determining the general number of clusters represented in a dataset with geometric coordinates.

**Why did you select that model?**

We used the model to identify trends in float glass location. We hypothesized that float glass plant locations will be clustered around areas of lower electricity price. The K-means clustering analyses supports this hypothesis, with electricity prices to float glass plants dense in the low electricity price range (4 to 8 cents per kilowatt-hour).

**How was the model trained**

The model includes convergence within its functionality, so it trains itself on identifying clusters by running the K-means clustering algorithm over dozens of iterations and on different cluster centers. The cluster center with the least distance to corresponding elements is eventually found and determines the origin of each cluster resulting from the model.

**Did you follow train, test and cross-validation protocols?**

We performed K-means clustering on 2 different cluster numbers: 3 and 4. For 3 clusters, we achieved a consistent result at 100 iterations that did not vary between runs. We achieved a consistent result for 4 clusters at 100 iterations as well. Validation was performed using the Elbow Method and the 3 cluster result was determined as the most accurate representation of clustering.

### **Communication**

**Have you annotated your notebook with clear explanations to walk the reader through, step by step?**

All operations are commented and any referenced code is included above each operation.

## **Data Analytics Final Project**

**Topic:** Analyzing Float Glass Plants by Electrical Rates

**Contributors:** Faris Alquaddoomi, Ethan Frisch, Megan Morrissey, Ajay Parthasarathy

### **Does your annotation communicate alternative paths considered in the analysis and decisions made for why particular directions were chosen?**

The rationale for each operation has been explained in the comments, as well as what the operation aims to achieve.

### **Have you clearly stated the conclusion from your project? Justify them**

We conclude that glass plant locations comprise 3 clusters: Western US, Midwest US, and Eastern US clusters. West and Midwest clusters are distributed within areas of predominantly lower electricity price, though we cannot say if this distribution is intentional.

We can charge an average of 6 cents per kilowatt-hour to remain competitive with on-grid electricity rates, with more or less charged depending on the exact location. This conclusion can be justified based on the first mode of the electricity price distribution, which is centered around 6 cents per kilowatt-hour.

### **What are future directions for the project were you to take more time and garner additional datasets.**

We want to generate an interactive web map of float glass plant locations and quantify the total revenue we can expect from each location. Revenue will be a function of the electricity price we charge and the carbon mitigation we can achieve (assuming carbon credits are included in our revenue).

We want to perform logistic regressions on carbon prices, using historical prices from Europe as training sets for our model. Europe is more advanced in implementing carbon credit programs and cap-and-trade programs, so they will serve as a valid training set for our predictive model. We want to predict future carbon credit prices in the US from the logistic regression.

We plan to save on carbon emissions with our waste heat recovery system and collect carbon credits per ton of CO<sub>2</sub> emitted. This is the second stream of revenue that we plan to incorporate as carbon credits become more integrated in the United States due to rejoining the Paris Climate Agreement. This will accompany the electricity savings at the glass float plants. We will also be able to predict carbon credit trends based on the already established system in Europe.

## **Data Analytics Final Project**

**Topic:** Analyzing Float Glass Plants by Electrical Rates

**Contributors:** Faris Alquaddoomi, Ethan Frisch, Megan Morrissey, Ajay Parthasarathy

### **Ethical Considerations**

**What ethical concerns, if any, may arrive as you consider your problem statement?**

- Incorrectly judging our electrical rate advantage over industrial rates could lead to our customers paying more than their best bet.
- TPVs Rare earth materials needed by mining: Gallium, Germanium
  - These materials are need by mining but in small amounts. As TPVs are scaled this will become a greater ethical concern

**Who might be affected?**

- Our customers
  - Overestimating will result in falsely optimistic forecasts on our revenue and could lead us to integrate more costly systems with our customers paying for the hardware upfront. This would predict an incorrect payback period on the customer's investment and consequently return to them less over the described period.
  - Underestimating will result in falsely pessimistic forecasts which would lead us to adopt a more risk-averse approach to implementing our system. We might implement a more cost-effective system that trades performance for a lower payback period as described to your customers. Our customer would then break even within a shorter period as described and generate less profit past break even.
- Our employees
  - Our employees stand to lose from incorrect predictions from our algorithms. Incorrect electrical rate advantage predictions would lower our profit, given our revenue is dependent on our customers' proven cost savings.
  - Lower revenue to our company from inaccurate customer cost savings would adversely affect our reputation and our ability to gain more clients. This would result in lower future profits and potentially the inability to pay salaries. In the worst-case scenario, we would have to terminate certain positions and layoff employees to meet our margin requirements.
- Electricity companies
  - By predicting what the electric input would be needed for the power plants where we implement our technology. If predicted wrong we might need to pay for more power than we are taking from the grid.

## **Data Analytics Final Project**

**Topic:** Analyzing Float Glass Plants by Electrical Rates

**Contributors:** Faris Alquaddoomi, Ethan Frisch, Megan Morrissey, Ajay Parthasarathy

### **Do your conclusions warrant any ethical considerations?**

Unrealistic optimism or pessimism in our forecasts negatively affects our company and our customers. If your conclusions are purely based on uncertain data, and we do not take the uncertainty into account when making future decisions, we will decrease the company profit, destroy the company reputation, fail to deliver returns to our stakeholders and fail to pay our employees the agreed-upon wages.

### **How might your model be abused by a bad-faith actor?**

Our data will not be abused. There is no incentive to use our predictions on electrical rates, as they would likely fit within the category of all predictions related to these parameters. As such, they would present no advantage to bad-faith actors for abuse beyond other predictions that exist and are based on more rigorous analyses.