**Instacart Online Shopping**

**Tools: Python, Jupyter Notebooks, Tableau**

**Libraries: Pandas, Numpy, Clustering, Geopandas, SKlearn Quandl.**

Posted by Graham Field on April 19, 2024 – 10 mins read

**Duration of the project:** 1 week

**Role:** Data Analyst

**Tools:**

* Python
* Jupyter Notebooks
* Libraries: Pandas, Numpy, Seaborn, Matplotlib

**Data:**

* Open Source Data
* > 3 000 000 Rows
* 23 Numeric and Categorical variables
* Joins: Order, Product, Customers and Department

**Key Research Questions**

* What are the busiest days of the week?
* When are the busiest hours of the day?
* Are there particular times of the day when spending is the highest?
* Who is placing the most orders?
* Which product departments are more popular than others?
* What is the distribution among users with regards to their brand loyalty and frequency of re-order and expenditure using mean, max and min?
* Is there a connection between age and family status in terms of ordering habits?

**Introduction**

The renewable energy industry has seen massive growth over the years given the importance of moving away from fossil fuels. One of the most effective ways of producing renewable energy is through wind. A global leader in this transition has been Germany.

The growth of the wind industry has many aspects to it, both technologically and geographically. With limited space and limited suitable areas for building wind farms the drive to make turbines more powerful and efficient is a major driving force in the industry.

Understanding how the new technologies have improved and how this has led to large increases in production capacity and physical dimensions of the turbines makes for an interesting study. Perhaps shedding light on the future of the industry.

**Step 1: Initial Data Exploration and Cleaning**

This is open-source data from the German Data Register (Marktstammdatenregister).

This data is regularly updated and free to use for research purposes.

**Step 2: Data Wrangling**

This is open-source data from the German Data Register (Marktstammdatenregister).

This data is regularly updated and free to use for research purposes.

**Step 3: Merging Data Sets**

This is open-source data from the German Data Register (Marktstammdatenregister).

This data is regularly updated and free to use for research purposes.

**Step 4: Deriving New Variables**

This is open-source data from the German Data Register (Marktstammdatenregister).

This data is regularly updated and free to use for research purposes.

**Step 5: Grouping Data**

This is open-source data from the German Data Register (Marktstammdatenregister).

This data is regularly updated and free to use for research purposes.

**Step 6: Visual Analysis**

This is open-source data from the German Data Register (Marktstammdatenregister).

This data is regularly updated and free to use for research purposes.

**Question 1**

What are the busiest days of the week?

**Question 2**

When are the busiest hours of the day?

**Question 3**

Are there particular times of the day when spending is the highest?

**Question 4**

Who is placing the most orders?

**Question 5**

Which product departments are more popular than others?

**Question 6**

What is the distribution among users with regards to their brand loyalty and frequency of re-order and expenditure using mean, max, and min?

**Question 7**

Is there a connection between age and family status in terms of ordering habits?