



# ELENA DRIES' DATA ANALYTICS PORTFOLIO

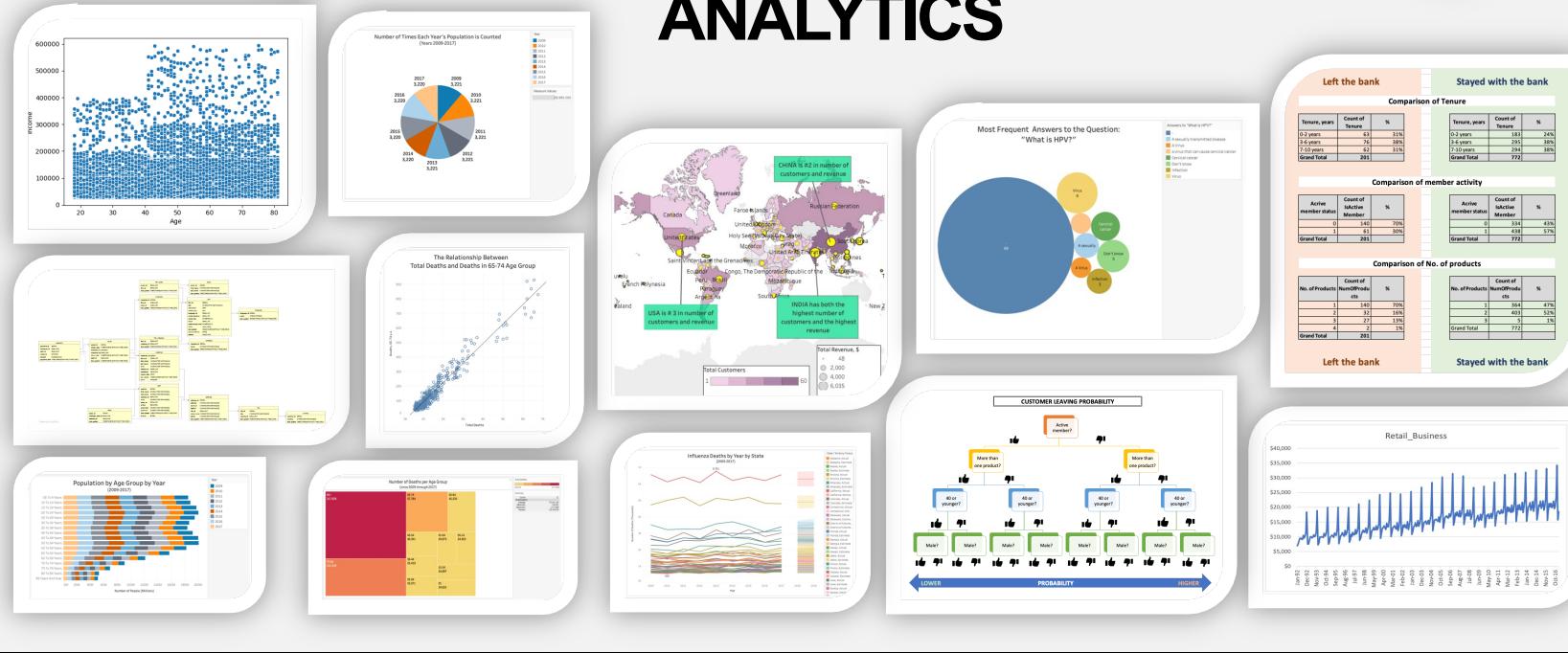
✉ [elena@dries.info](mailto:elena@dries.info)  
⌚ [www.ElenaDries.com](http://www.ElenaDries.com)



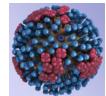
# TECH STACK

---

# ANALYTICS



# PROJECT LIST



1. Preparing for Influenza Season



2. Rockbuster's Top Performers



3. Instacart Grocery Basket Analysis

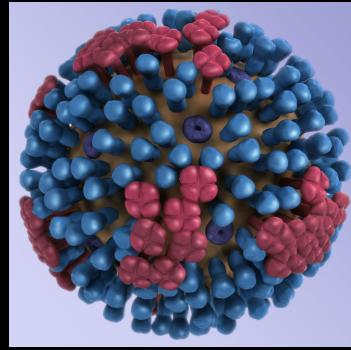


4. US Residential Real Estate



1.

# PREPARING FOR INFLUENZA SEASON



# PREPARING FOR INFLUENZA SEASON PROJECT OVERVIEW

## Project Goals

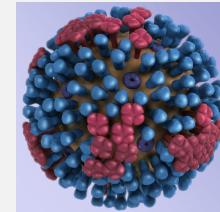
- Determine how much additional temporary staff to provide to adequately treat patients during flu season in the US.
- Distribute the staff across different states.

## Tools/Techniques

- Excel
- Translating business requirements
- Data cleaning/integration/transformation
- Statistical hypothesis testing
- Visual analysis
- Forecasting
- Storytelling in Tableau

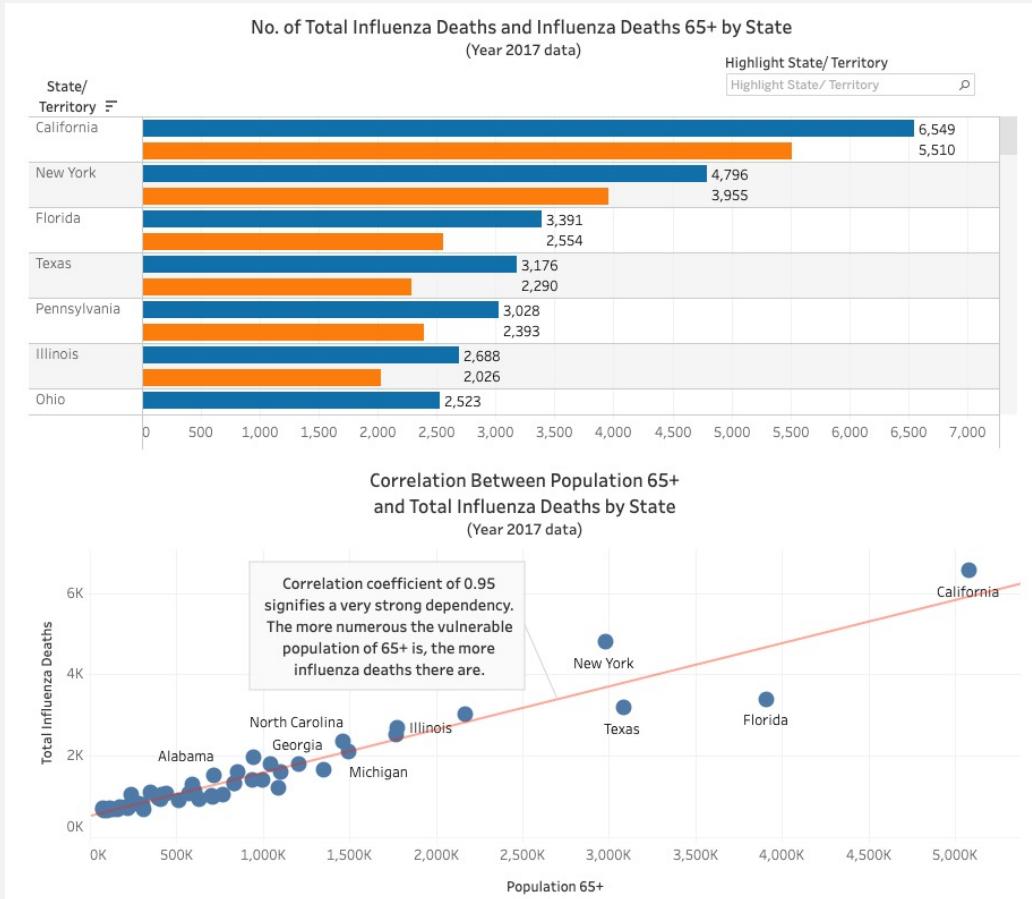
## Data

- Influenza deaths by geography, time, age, and gender (Source: CDC)  
[https://coach-courses-us.s3.amazonaws.com/public/courses/d\\_a\\_program/CDC\\_Influenza\\_Deaths\\_edited.xlsx](https://coach-courses-us.s3.amazonaws.com/public/courses/d_a_program/CDC_Influenza_Deaths_edited.xlsx)
- Population data by geography (Source: US Census Bureau)  
[https://coach-courses-us.s3.amazonaws.com/public/courses/data-immersion/A1-A2\\_Influenza\\_Project/Census\\_Population\\_transformed\\_202101.csv](https://coach-courses-us.s3.amazonaws.com/public/courses/data-immersion/A1-A2_Influenza_Project/Census_Population_transformed_202101.csv)
- Influenza visits data set (Source: CDC (Fluvview))  
[https://images.careerfoundry.com/public/courses/data-immersion/A1-A2\\_Influenza\\_Project/CDC\\_Influenza\\_Visits.xlsx](https://images.careerfoundry.com/public/courses/data-immersion/A1-A2_Influenza_Project/CDC_Influenza_Visits.xlsx)



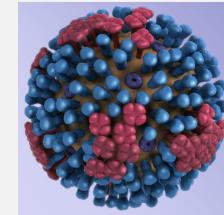
# PREPARING FOR INFLUENZA SEASON

## CHARTS / INTERIM POINTS



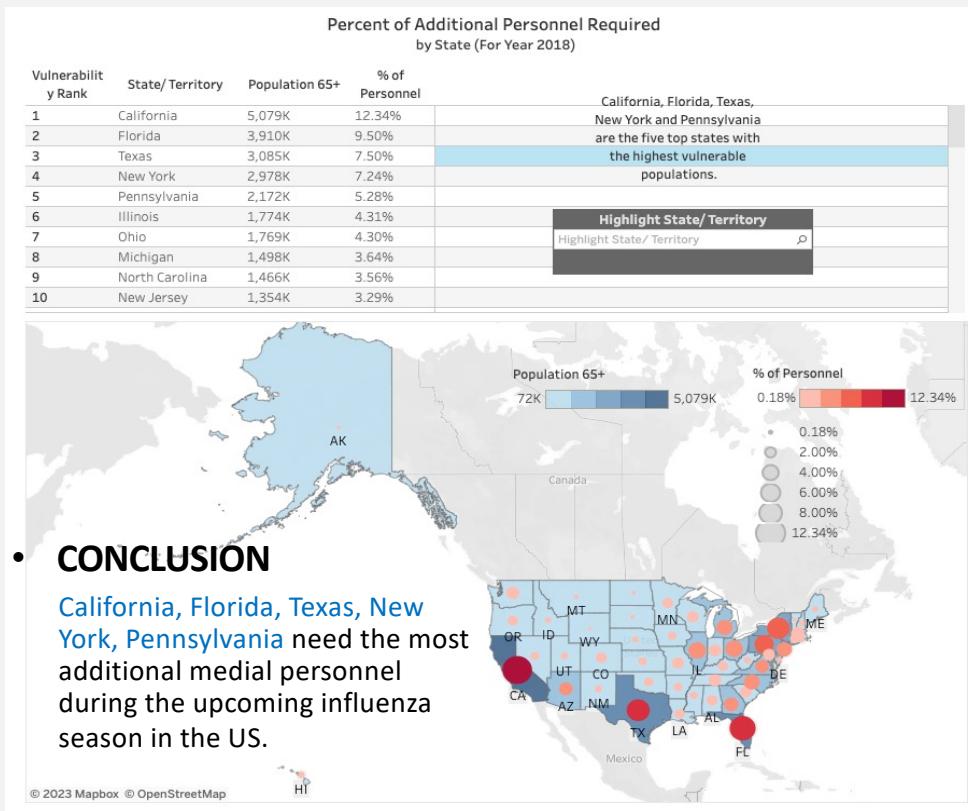
## Interim Points

- Flu deaths are largely driven by the number of vulnerable population of 65+
- The larger the vulnerable population, the greater the need for additional staff



# PREPARING FOR INFLUENZA SEASON RESULTS AND DELIVERABLES

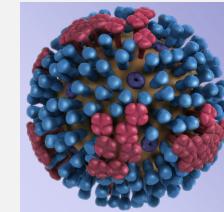
## Results



## Deliverables

- YouTube Video Presentation:**  
<https://youtu.be/l6eAgzEQmwg>

- Tableau Presentation (with interactive charts):**  
[https://public.tableau.com/app/profile/elenadries/viz/2\\_9\\_FinalProjectbyED/InfluenzaAdditionalPersonnelSupport?publish=yes](https://public.tableau.com/app/profile/elenadries/viz/2_9_FinalProjectbyED/InfluenzaAdditionalPersonnelSupport?publish=yes)



2.

## ROCKBUSTER'S TOP PERFORMERS



# ROCKBUSTER'S TOP PERFORMERS PROJECT OVERVIEW



## Project Goals

- Where are Rockbuster's customers based?
- How different are the sales in different countries?
- Who are the top 5 customers for recognition?

## Tools/Techniques

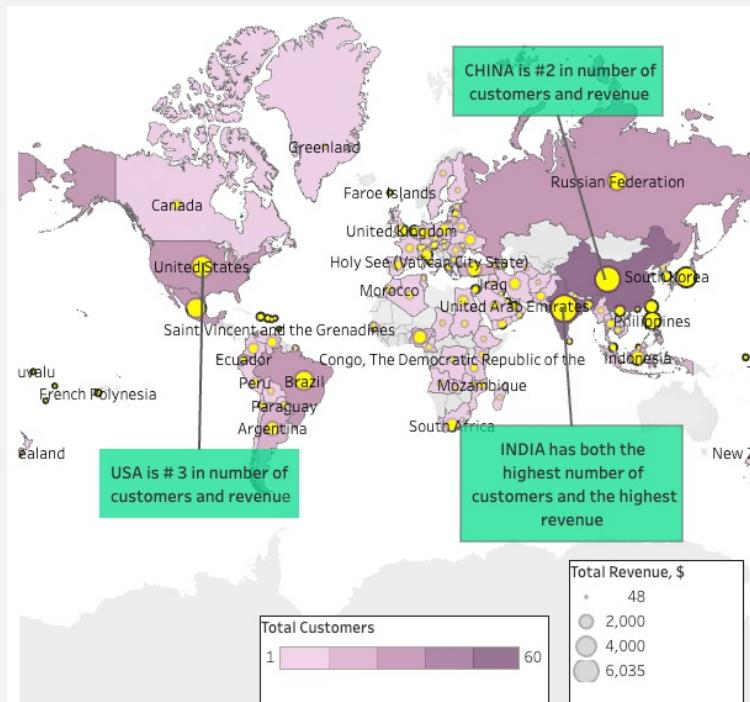
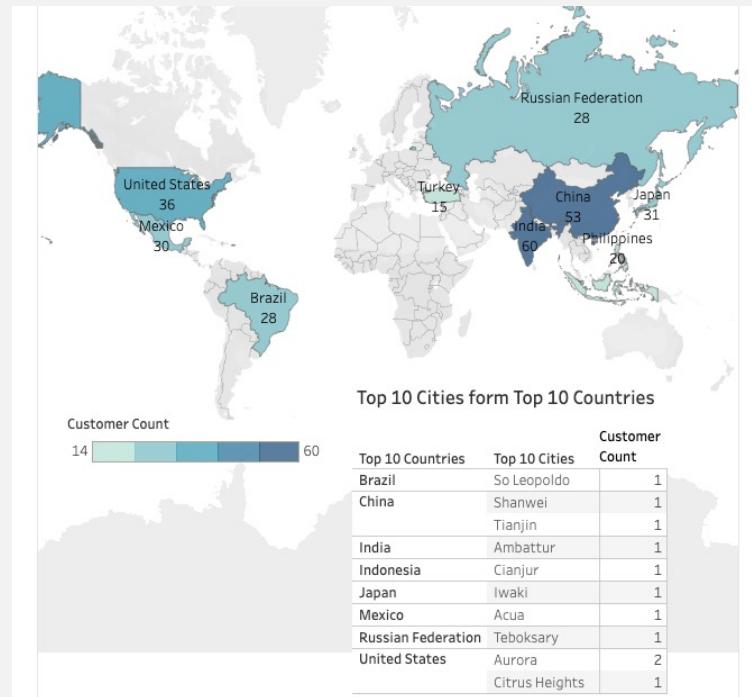
- Relational databases
- SQL
- Filtering, cleaning, summarizing
- Joining tables
- Subqueries
- Common table expressions

## Data

- Database backup file (.zip):  
[https://careerfoundry.com/en/course/data-immersion/exercise/intro-to-relational-databases#relational-database-management-systems~:text=download%20the%20database%20backup%20file%20\(.zip\)](https://careerfoundry.com/en/course/data-immersion/exercise/intro-to-relational-databases#relational-database-management-systems~:text=download%20the%20database%20backup%20file%20(.zip))



# ROCKBUSTER'S TOP PERFORMERS CHARTS / INTERIM POINTS



- Here is the list of the top 10 cities from the top 10 countries by the number of customers

**Interim Points**

- India, China and the USA are leading in the geographical distribution of customers and generated revenues



Elena Dries

10

# ROCKBUSTER'S TOP PERFORMERS RESULTS AND DELIVERABLES



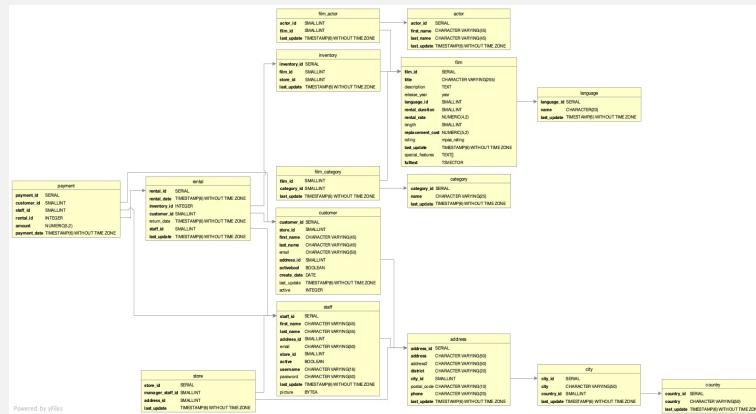
## Results

Customer ID	Customer First Name	Customer Last Name	Country	City	Total Amount Paid, \$
225	Arlene	Harvey	India	Ambattur	\$111.76
424	Kyle	Spurlock	China	Shanwei	\$109.71
240	Marlene	Welch	Japan	Iwaki	\$106.77
486	Glen	Talbert	Mexico	Acua	\$100.77
537	Clinton	Buford	United States	Aurora	\$98.76

## CONCLUSION

Here are the top 5 customers in the top 10 cities, who paid the highest total amounts to Rockbuster.

## Entity Relationship Diagram (ERD):



## Deliverables

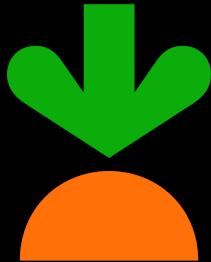
- GitHub Repository:  
<https://github.com/Elena3s/RockbustersTopPerformers>

- Tableau Presentation (with interactive charts):  
[https://public.tableau.com/app/profile/elenadries/viz/3\\_10byElenaDries/Story1?publish=yes](https://public.tableau.com/app/profile/elenadries/viz/3_10byElenaDries/Story1?publish=yes)



Elena Dries

11



# 3. INSTACART GROCERY BASKET ANALYSIS



# INSTACART GROCERY BASKET ANALYSIS

## PROJECT OVERVIEW



### Project Goals

- Uncover sales patterns
- Derive insights on better market segmentation

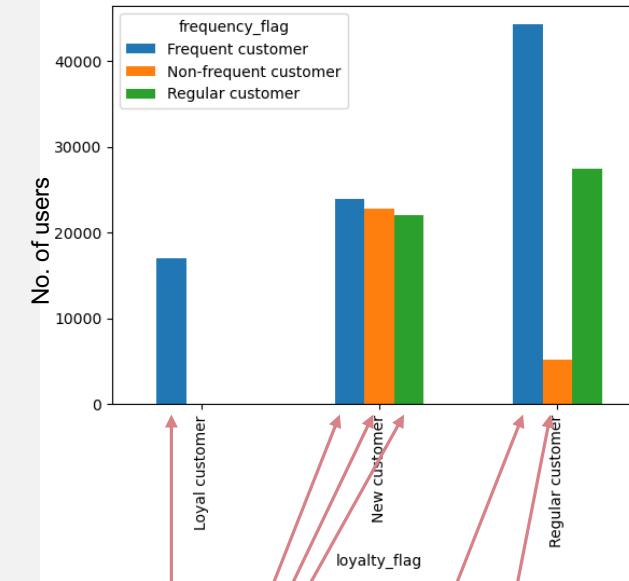
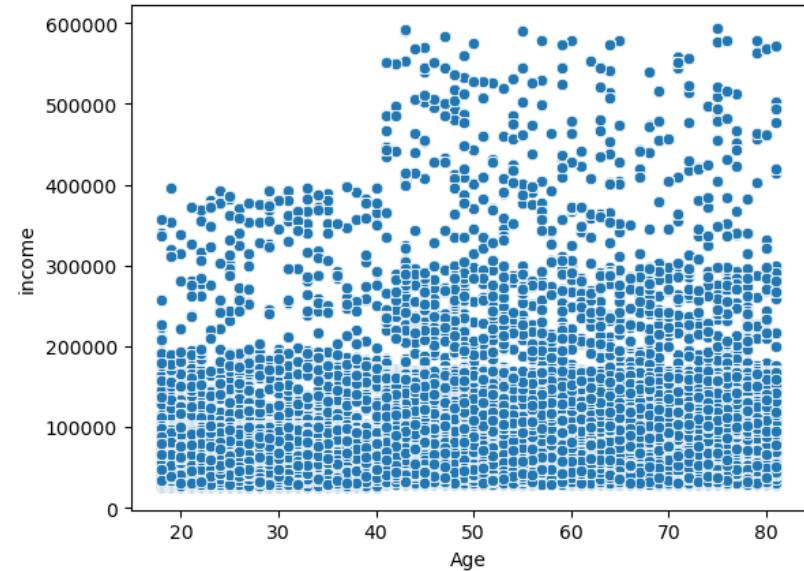
### Tools/Techniques

- Python
- Pandas
- Data wrangling, merging
- Deriving variables
- Grouping and aggregating data
- Reporting in Excel
- Population flows

### Data

- The Instacart Online Grocery Shopping Dataset 2017, accessed in December 2022:  
<https://www.instacart.com/datasets/grocery-shopping-2017>
- Orders and Products (.zip):  
[https://s3.amazonaws.com/coach-courses-us/public/courses/data-immersion/A4/A4\\_Data\\_Assets/4.3\\_orders\\_products.zip](https://s3.amazonaws.com/coach-courses-us/public/courses/data-immersion/A4/A4_Data_Assets/4.3_orders_products.zip)
- Departments (.zip):  
[https://s3.amazonaws.com/coach-courses-us/public/courses/data-immersion/A4/A4\\_Data\\_Assets/4.4\\_departments.zip](https://s3.amazonaws.com/coach-courses-us/public/courses/data-immersion/A4/A4_Data_Assets/4.4_departments.zip)
- Customers (.zip):  
[https://s3.amazonaws.com/coach-courses-us/public/courses/data-immersion/A4/A4\\_Data\\_Assets/customers.zip](https://s3.amazonaws.com/coach-courses-us/public/courses/data-immersion/A4/A4_Data_Assets/customers.zip)
- Data Dictionary:  
<https://gist.github.com/jeremystan/c3b39d947d9b88b3ccff3147dbcf6c6b>

# INSTACART GROCERY BASKET ANALYSIS CHARTS / INTERIM POINTS

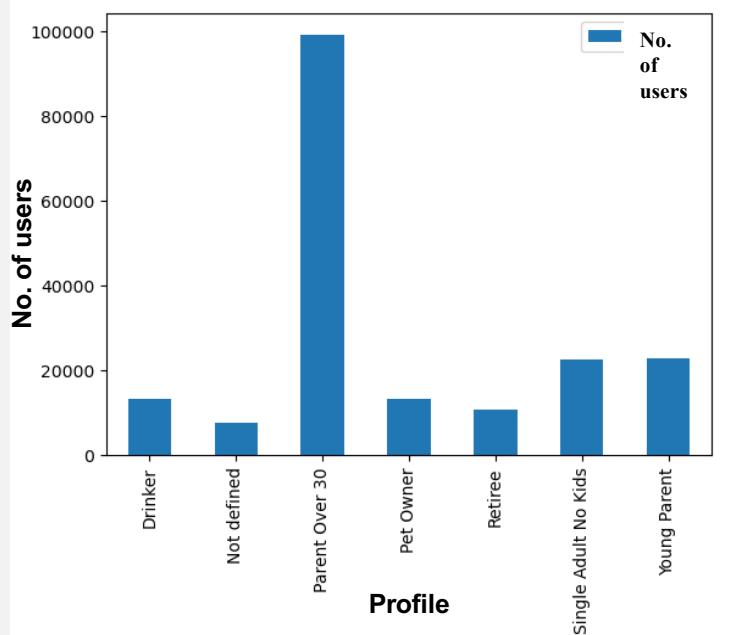


- At the age of 40 many Instacart customers transition from <200K to <300K income. The income ceiling also jumps up from 400K to 600K at the same age line of 40.
- The most loyal customers (with over 40 orders) are also frequent customers ( $\leq 10$  days between orders)
- The regular customers in terms of loyalty (from 11 to 40 orders) show a prevalence of being frequent in their purchases, while only a small portion of them have more than 20 days between their orders.
- The new customers are evenly distributed among the frequency categories.

# INSTACART GROCERY BASKET ANALYSIS

## RESULTS AND DELIVERABLES

### Results



- The profile category "Parent over 30" (and below 65, with dependents) is by far the most numerous one. (NB: Some profiles may overlap.)

Region	High spender	Low spender
Region 1	108,225	5,614,511
Region 2	155,975	7,441,350
Region 3	209,691	10,582,194
Region 4	160,354	8,132,559

- Region 3 has the most customers and almost doubles both High and Low spenders when compared to Region 1. Regions 2 and 4 are hanging somewhere in between.

### Deliverables

- GitHub Repository:**  
<https://github.com/Elena3s/InstacartGroceryBasketAnalysis>



4.

# US RESIDENTIAL REAL ESTATE



# US RESIDENTIAL REAL ESTATE ANALYSIS PROJECT OVERVIEW



## Project Goals

- Investigate how/if the property size affects its sq.ft. price (2016-2023)
- Investigate the relationship between property size and how fast it closes (2016-2023)
- Discover the hottest residential markets as of 2022

## Tools/Techniques

- Python
- Pandas/seaborn/  
matplotlib/sklearn
- Linear regression
- Elbow technique
- k-means clustering
- Time series

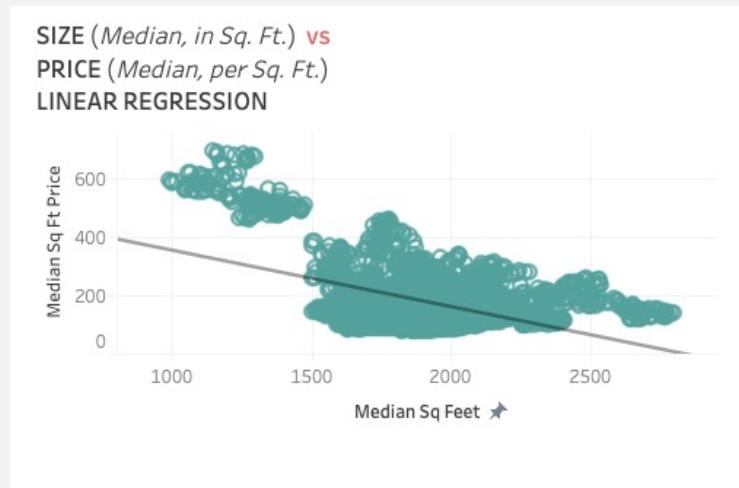
## Data

- Publicly available source on realtor.com:  
[https://www.realtor.com/  
research/data/](https://www.realtor.com/research/data/)

*(downloaded in February 2023, updated on a monthly basis)*

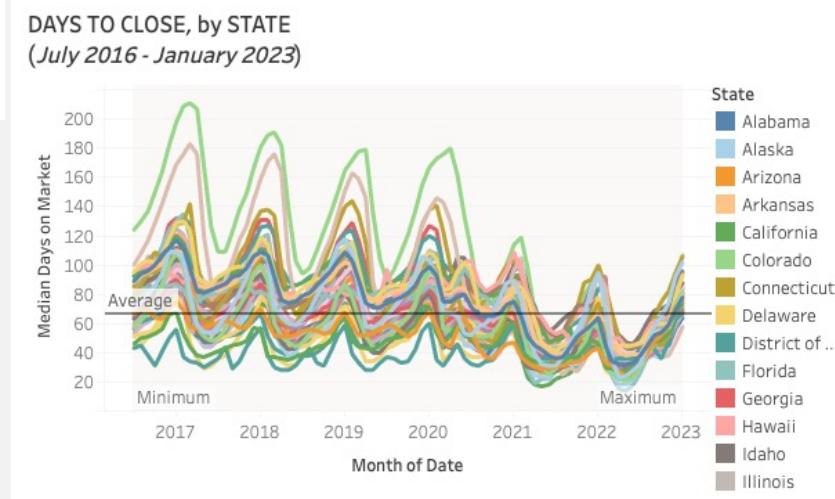


# US RESIDENTIAL REAL ESTATE ANALYSIS CHARTS / INTERIM POINTS



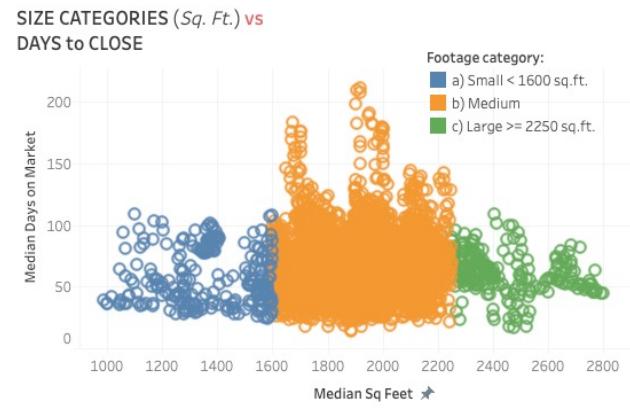
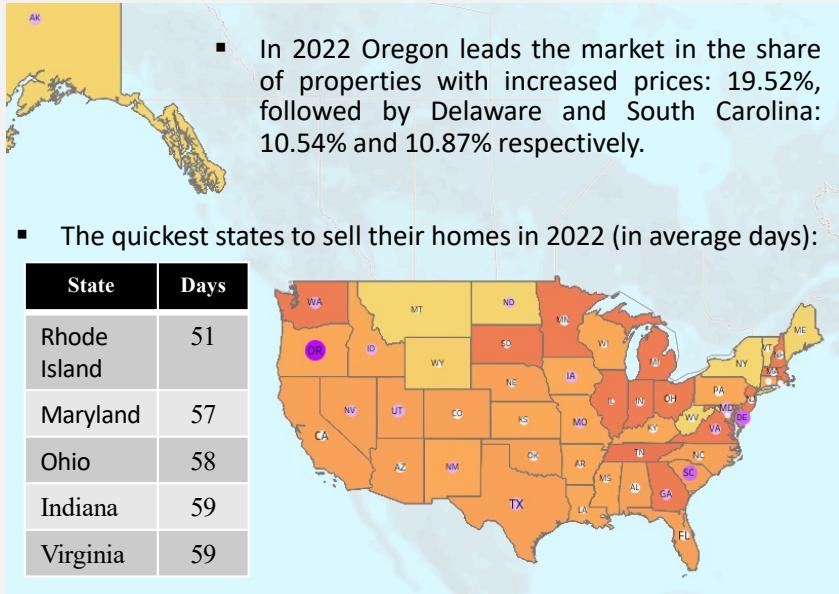
- The property size contributes about 48% to its sq.ft. price - medium strength dependency, which means that even though it plays a considerable role in defining the square foot price we cannot fully rely on the property size in order to explain the behavior of its square foot price.

- The national average number of days to close between 2016-2023 is at 66.8 days. The number one state with the shortest average time to sell is the District of Columbia - at 40.3. The slowest' is Vermont - at 118.7 days.



# US RESIDENTIAL REAL ESTATE ANALYSIS RESULTS AND DELIVERABLES

## Results

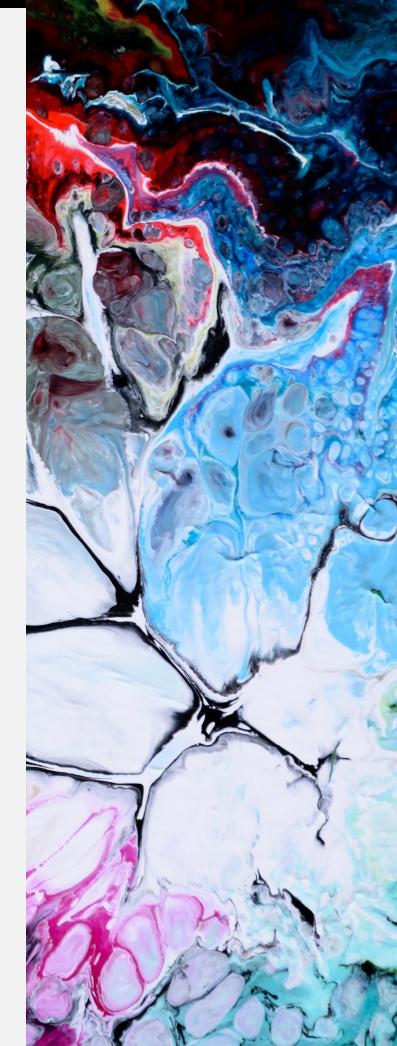


- Small and large properties, on average, have very little chance to get stuck on the market for longer than 110 days.
- Medium-sized properties have a risk of hanging there for much longer - up to 210 days, even though the majority of them do sell in under 110 days.

## Deliverables

- GitHub Repository:  
<https://github.com/Elena3s/US-Residential-RE-Analysis>

- Tableau Presentation (with interactive charts):  
[https://public.tableau.com/app/profile/elena.dries/viz/USRealEstateMarket\\_16800293963450/USREMarket?publish=yes](https://public.tableau.com/app/profile/elena.dries/viz/USRealEstateMarket_16800293963450/USREMarket?publish=yes)



# CONTACT INFO

✉ [elena@dries.info](mailto:elena@dries.info)  
🌐 [www.ElenaDries.com](http://www.ElenaDries.com)

# SOCIAL MEDIA

