

Instacart



PROBLEM:

- Instacart is considering a targeted marketing strategy. In order to target the right customer profiles with the appropriate products, they want to learn more about their sales trends and purchasing behaviour.

OBJECTIVE:

- Perform an initial data and exploratory analysis of some of Instacart data in order to derive insights and suggest strategies for better segmentation.

KEYNOTES

- I conducted customer profiling based on the customer's age, income, number of dependents, and whether they were purchasing things for babies, pets, or vegetarians.
- Based on client loyalty, purchasing patterns, and other factors, I generated additional columns (flags).
- When comparing two flags, the Panda's crosstab function was quite beneficial.

SKILLS

- Data wrangling
- Data merging
- Deriving variables
- Grouping and aggregating the data
- Reporting in Excel
- Population flows

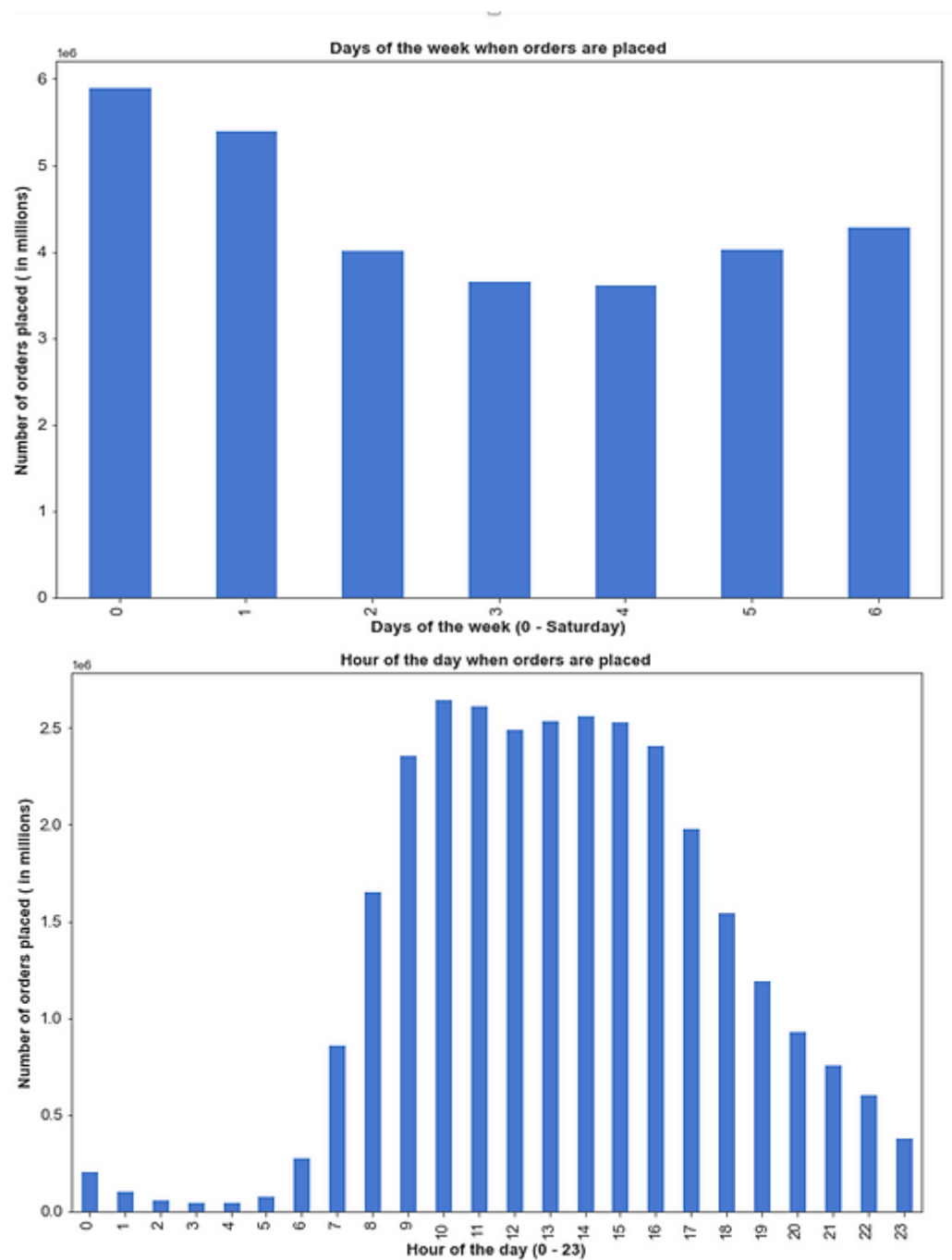
TOOLS

- Python
- Jupyter
- Excel

DATA

- [Orders](#)
- [Products](#)
- [Departments](#)
- [Customers](#)

What are the busiest days of the week and hours of the day?



- The busiest days are Saturday and Sunday, this tells us that people love to buy groceries on the weekend when they have more free time.
- The busiest hours are between 9 and 16 h.

Are there differences in ordering habits based on a customer's region

Defining Regions

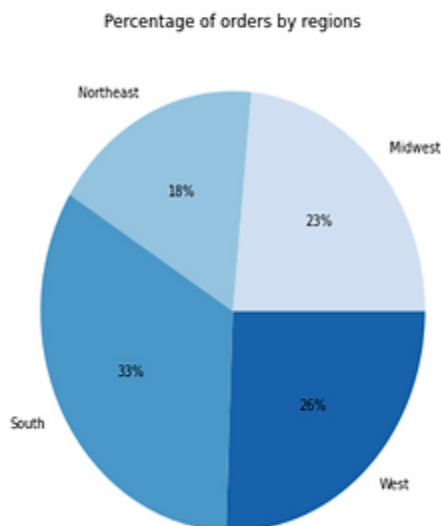
```
# create a regional segmentation of the data
regions = {'Maine' : 'Northeast', 'New Hampshire' : 'Northeast', 'Vermont' : 'Northeast', 'Massachusetts' : 'Northeast', 'Wisconsin' : 'Midwest', 'Michigan' : 'Midwest', 'Illinois' : 'Midwest', 'Indiana' : 'Midwest', 'Ohio' : 'Midwest', 'Delaware' : 'South', 'Maryland' : 'South', 'District of Columbia' : 'South', 'Virginia' : 'South', 'West Virginia' : 'South', 'Idaho' : 'West', 'Montana' : 'West', 'Wyoming' : 'West', 'Nevada' : 'West', 'Utah' : 'West', 'Colorado' : 'West'}

# define function get_region
def get_region(state):
    return regions[state]

cust_ords_prods_merge['region'] = cust_ords_prods_merge['state'].apply(get_region)

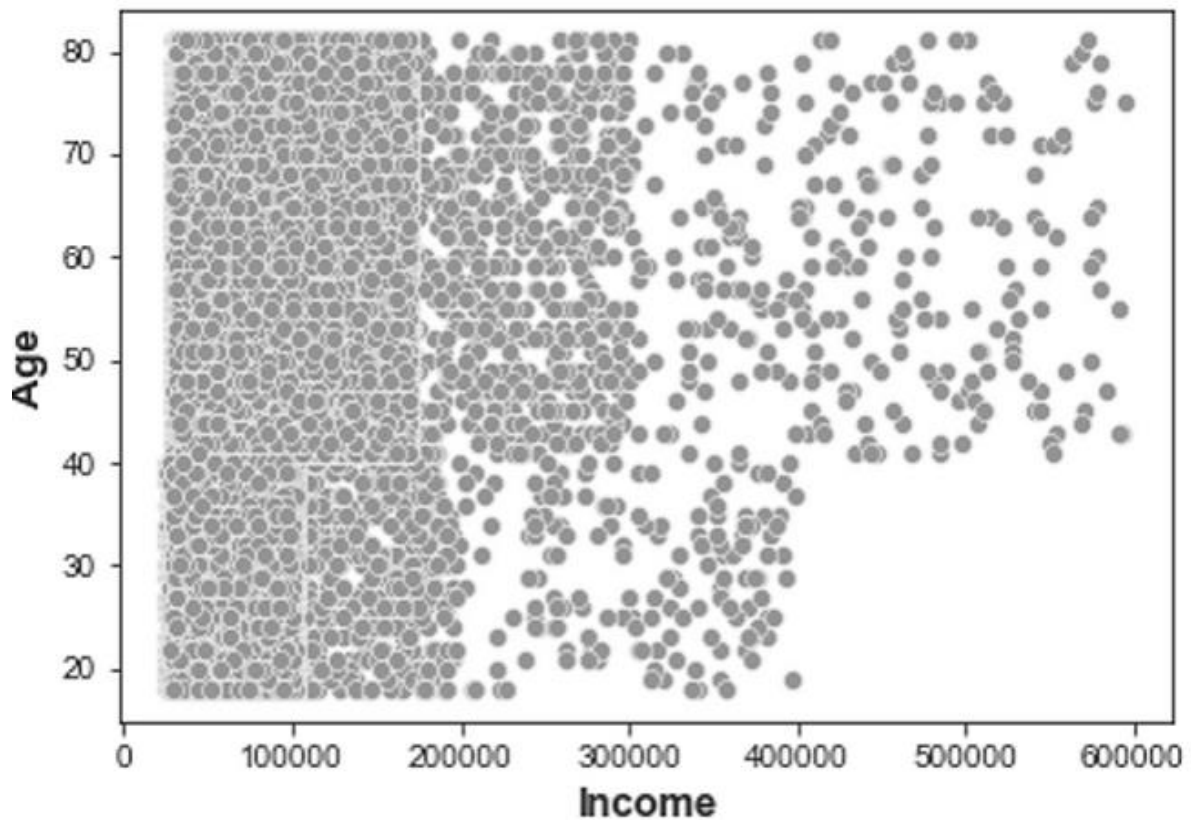
data = cust_ords_prods_merge[['region', 'order_id']].groupby(['region']).nunique().sort_index()

region_orders_pie = data['order_id'].plot.pie(figsize = (7,8), colors = sns.color_palette('Blues', 4), autopct = '%2.1f%%')
plt.ylabel('')
plt.title('Percentage of orders by regions', fontsize=13)
plt.show()
```



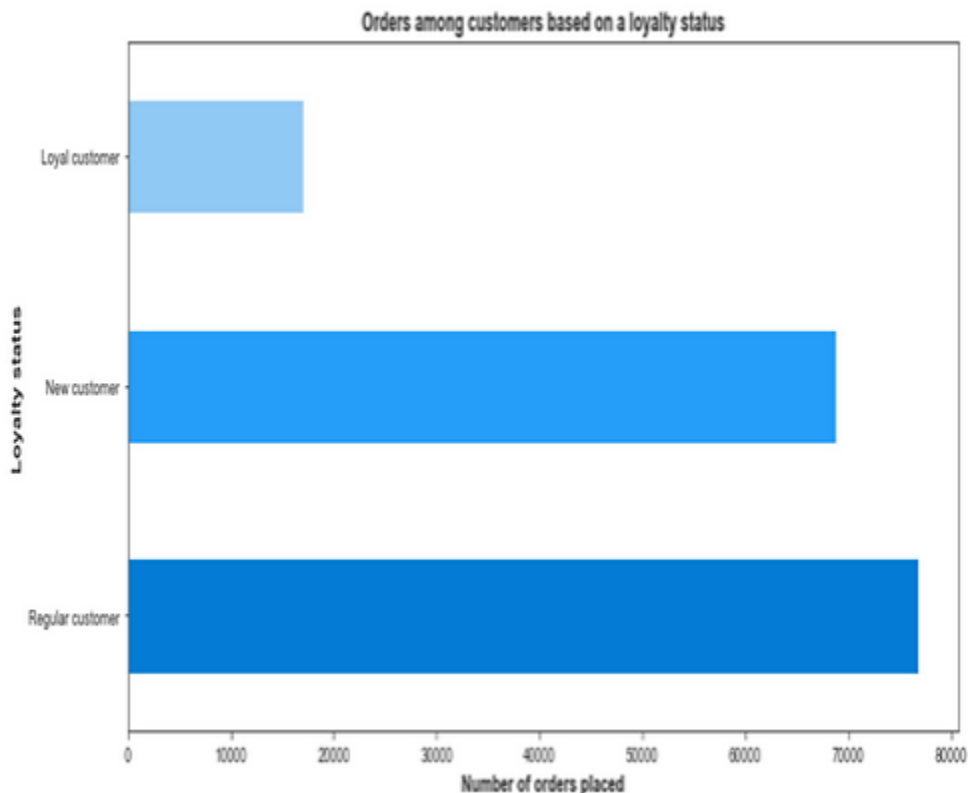
- I created a regional data segmentation to compare customer behavior across different regions.
- The South region has the biggest number of customers, corresponding with the highest number of orders being made in the same region.

What different classifications does the demographic information suggest? Age? Income?



- Most customers have equal purchasing power (up to \$200.000).
- Customers above the age of 40 have larger purchasing power (incomes greater than \$400.000).

Are there differences in ordering habits based on a customer's loyalty status?



- Instacart has the majority of orders from regular customers, which is followed by new customers. The number of orders that are made from loyal customers suggests that IC needs to do something to make Regular and new customers come back more often.

Creating Flags

```
# create a flag that assigns a "loyalty" label to a user ID based on its corresponding max order value
ords_prods_merge.loc[ords_prods_merge['max_order'] > 40, 'loyalty_flag'] = 'Loyal customer'

ords_prods_merge.loc[(ords_prods_merge['max_order'] <= 40) & (ords_prods_merge['max_order'] > 10), 'loyalty_flag']

ords_prods_merge.loc[ords_prods_merge['max_order'] <= 10, 'loyalty_flag'] = 'New customer'

ords_prods_merge['loyalty_flag'].value_counts(dropna = False)

Regular customer    15876776
Loyal customer      10284093
New customer        6243990
Name: loyalty_flag, dtype: int64
```

A customer is considered to be "Loyal" if they have placed over 40 orders, "Regular" if they have placed more than 10 and less than or equal to 40 orders, and "New" if they have placed fewer than 10 orders overall

RECOMMENDATIONS

- Advertise more during the weekdays in the early mornings and late afternoon.
- Concentrate more on the Northeast and Midwest regions.

- Reward Loyalty with special offers to loyal customers.
- IC should target more older customers because they have higher purchasing.