Data Analyst Professional Practical Exam Presentation

Dr. Anoushiravan zahedi

Pens and Printer

- Founded in 1984 and provides high-quality office products
- New Products
- Sales Methods
  - Email
  - Call
  - Email and Call

Data Validation

Columns

Column Name	Details	Required Modifications
week	Week sale was made, counted as weeks since product launch	NA
customer_id	Character, unique identifier for the customer	Transformed to Category
nb_sold	Numeric, number of new products sold	NA
years_as_customer	years_as_customer Numeric, number of years customer has been buying from us(company founded in 1984)	Out of range values replaced with NaN
nb_site_visits	Numeric, number of times the customer has visited our website in the last 6 months	NA
state	Character, location of the customer i.e. where orders are shipped	Transformed to Category
sales_method	Character, which of the three sales methods were used for that custome	Cleaned & Transformed to Category
revenue	Numeric, revenue from the sales, rounded to 2 decimal places.	Missing data

Sales Method

```
print(f"value counts for sales_method: \n{data['sales_method'].value_counts()}")
```

Sales Method	Count
Email	7456
Call	4962
Email + Call	2549
em + call	23
email	10

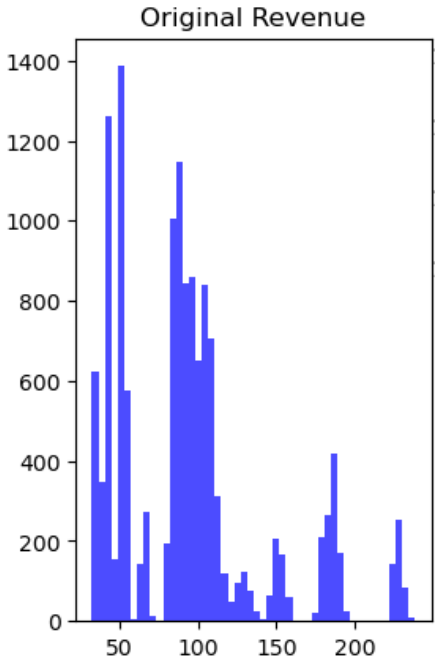
```
data['sales_method'] = data['sales_method'].str.replace('em ','email ').str.title().astype('category')
```

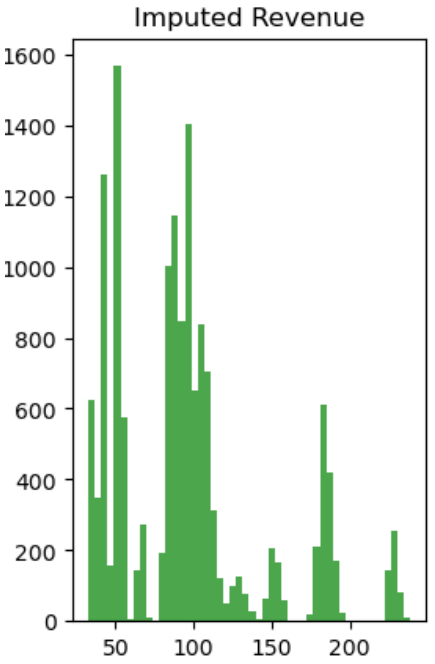
Revenue

```
print(f"number of na in revenue grouped by sales_method: \
{data.groupby('sales_method', observed=False)['revenue'].apply(lambda x: x.isna().sum()/len(x))}")
```

Sales Method	Proportion of NA
Call	0.036477
Email	0.072864
Email + Call	0.135692

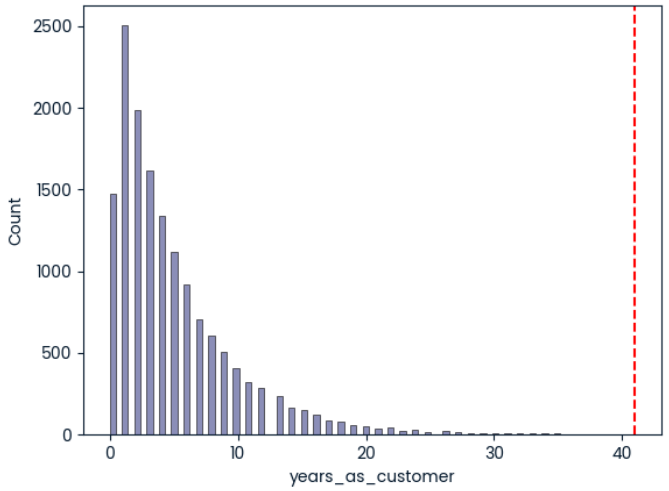
```
data['revenue_filled'] = data.groupby('sales_method', observed=False)['revenue'].\
transform(lambda x: x.fillna(x.median()))
```

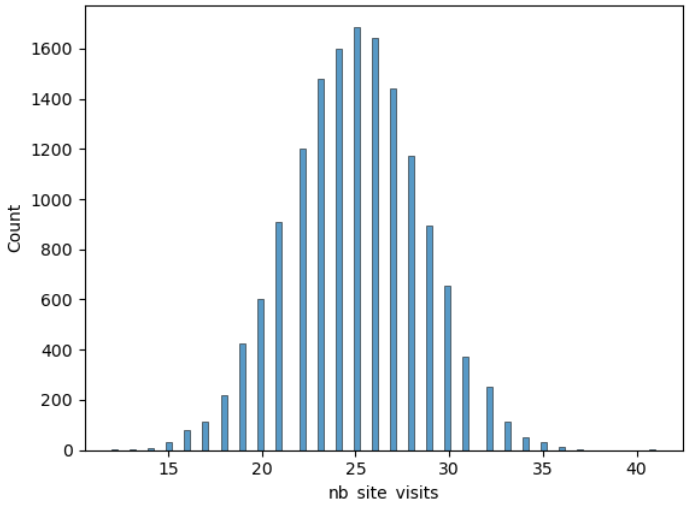


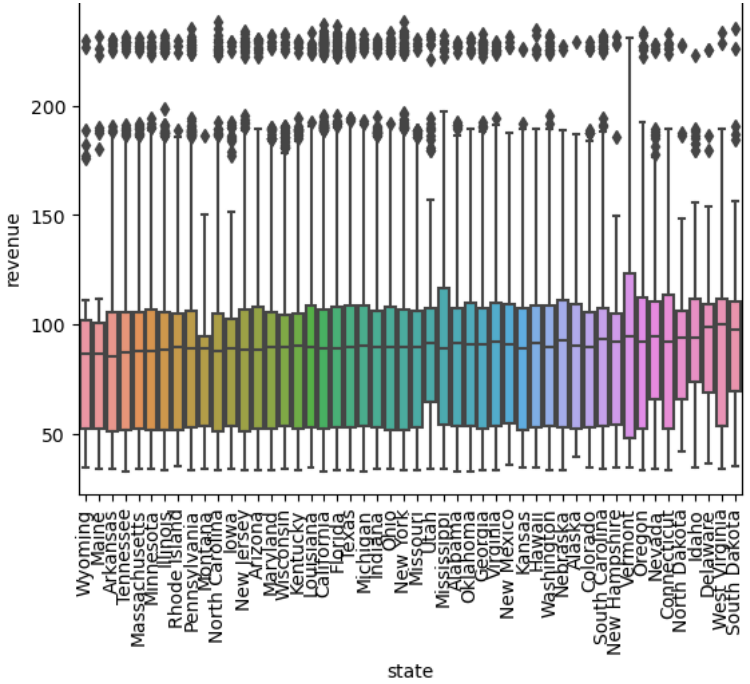


Exploratory analysis

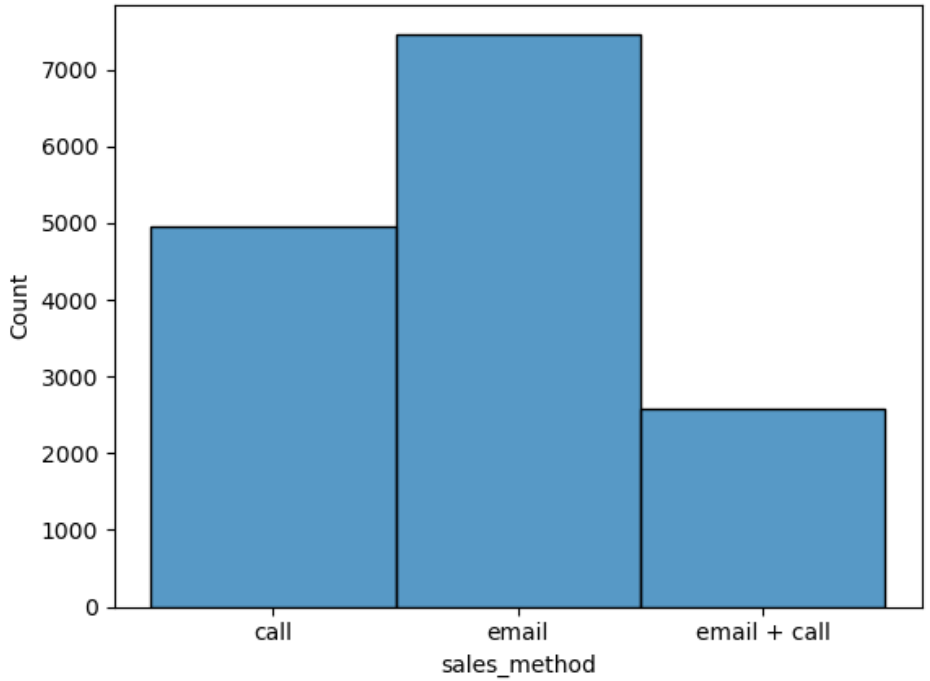
Customers and their characteristics



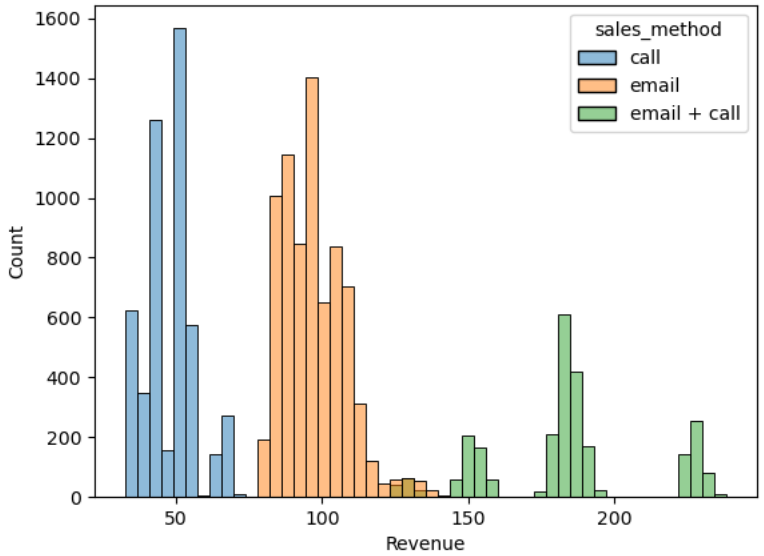


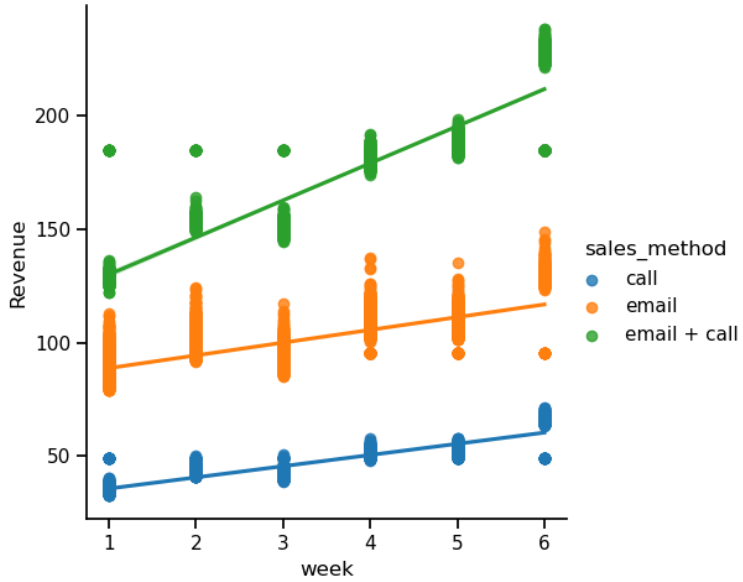


Allocation of customers to sales methods



Revenue for different groups





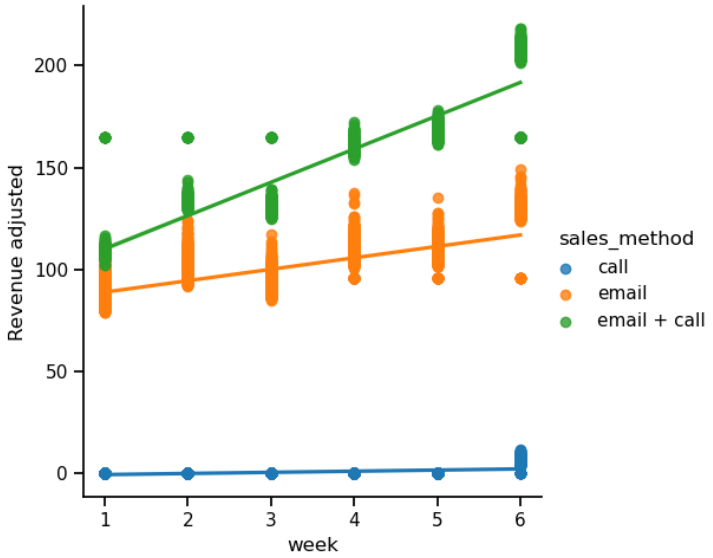
Contrast	A	B	U-val	p-corr	hedges
sales_method	email	email + call	15854.5	0.0	-5.241444
sales_method	email	call	37046292.0	0.0	4.969941
sales_method	email + call	call	12762264.0	0.0	7.903535

Definition of a metric for the business

Adjusted revenue

```
## finding a relevant metric
conditions = [
    data['sales_method'] == 'Email',
    data['sales_method'] == 'Call',
    data['sales_method'] == 'Email + Call',
]
choice = [0,30,10]
data['time_used'] = np.select(conditions, choice)
```

```
time_cost=2
data['revenue_adj'] = np.maximum(data['revenue_filled']-data['time_used']*time_cost,0)
```



A	B	U-val	p-corr	hedges
email	email + call	15854.5	0.0	-5.241444
email	call	37046292.0	0.0	4.969941
email + call	call	12762264.0	0.0	7.903535

Conclusions

1. Email & call > Email > Call

2. The differences between the sales methods are significant.

3. Using adjusted revenue

Thank you

Dr. Anoushiravan Zahedi

anoushiravanzahedi@gmail.com