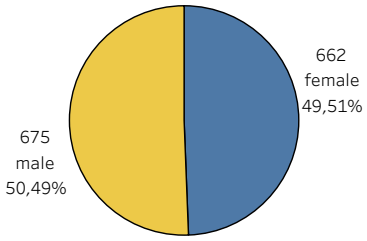
Solutions: customer acquisition campaign promoting health for ACME INSURANCE

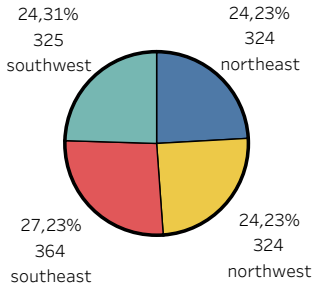
May 2023

Soft data

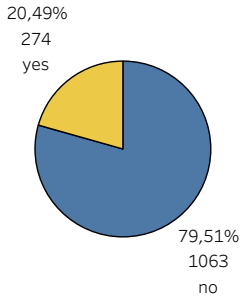
Sex



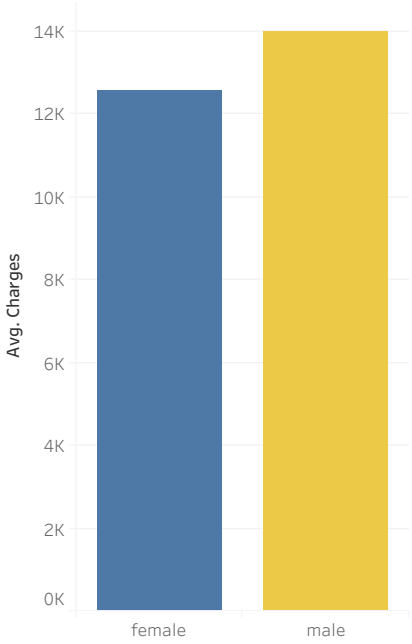
Region



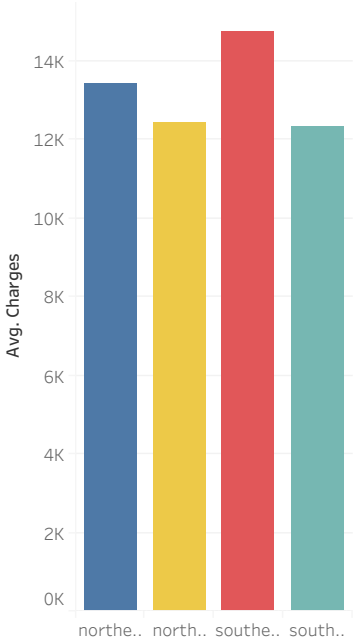
Smoker status



Charges by Sex

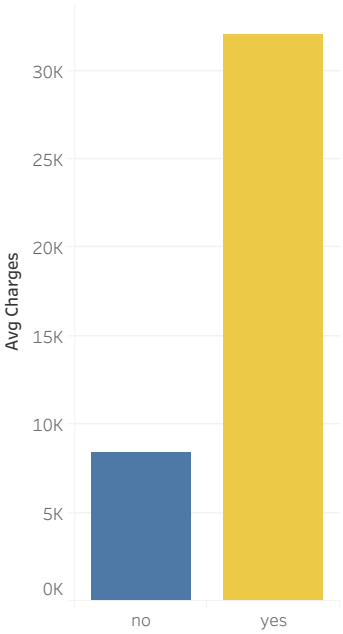


Charges by Region

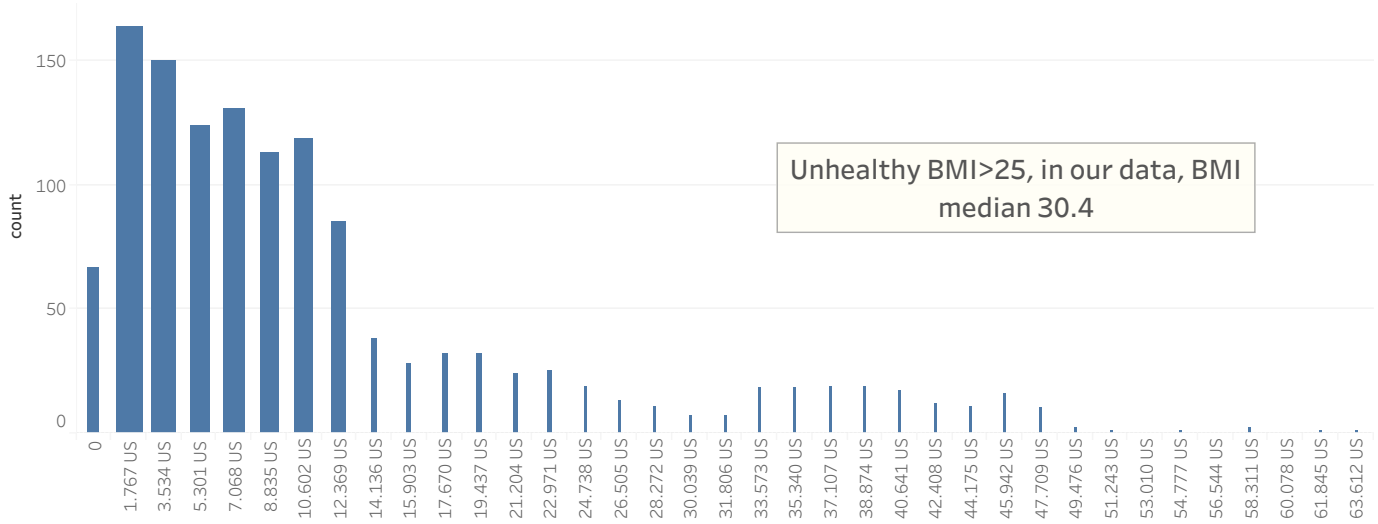


Ammount insurance premiums suffer a dramatic increase to double on average for smokers

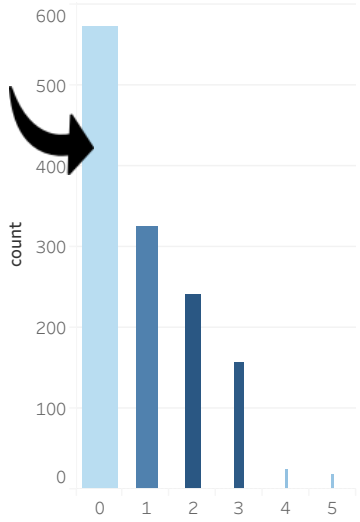
Charges by Smoker status



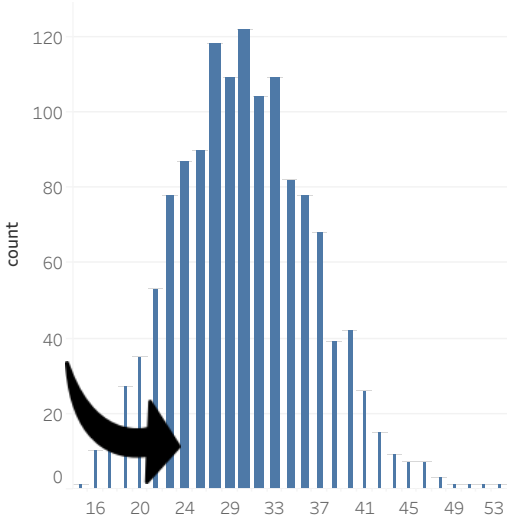
Charges frecuency-distribution



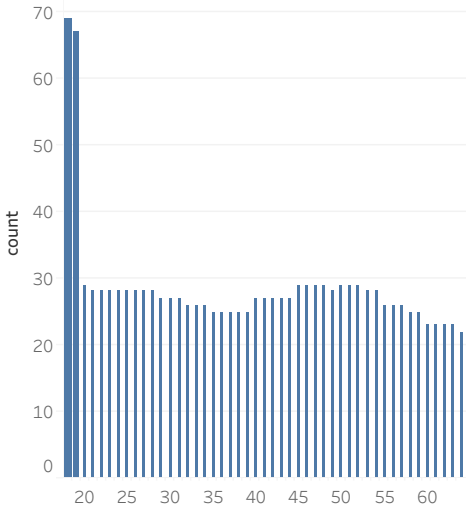
Children



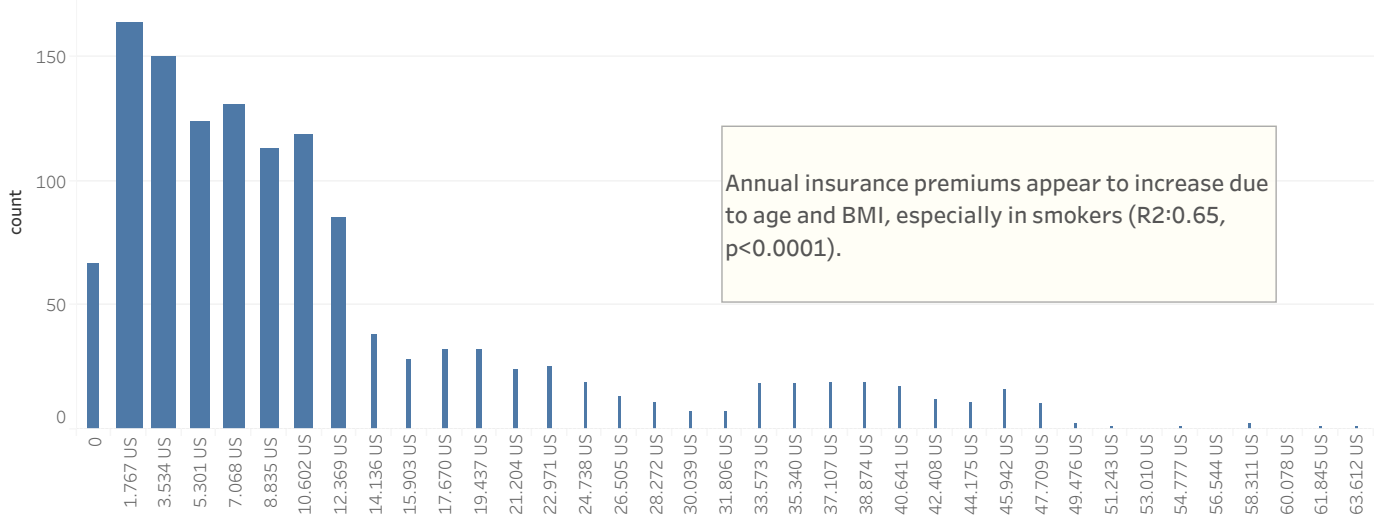
BMI frecuency-distribution



Age frecuency

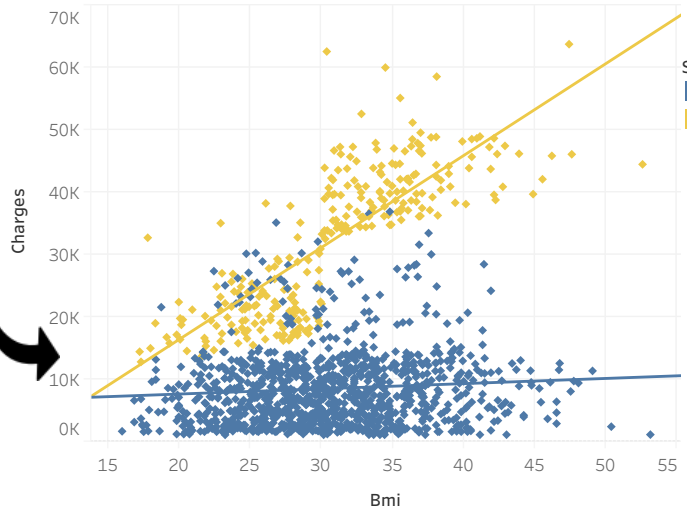


Charges frecuency-distribution

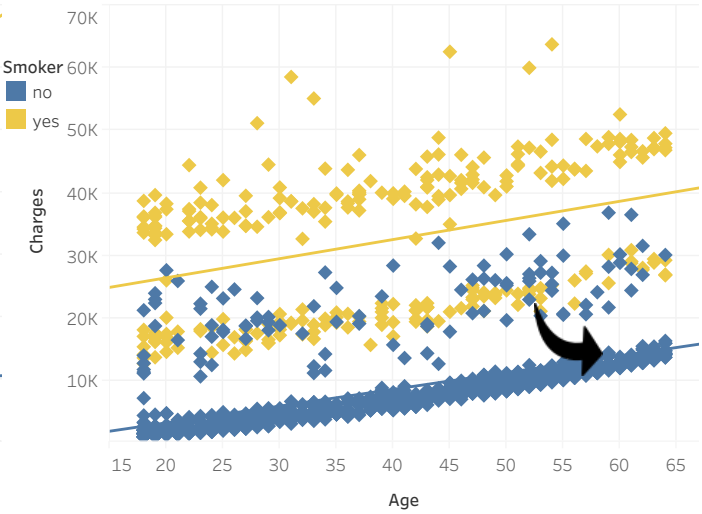


Annual insurance premiums appear to increase due to age and BMI, especially in smokers (R2:0.65, p<0.0001).

Charges-BMI by smoker status



Charges-Age by smoker status



100%| 42/42 [00:33<00:00, 1.24it/s]

Model	Adjusted R-Squared	R-Squared	RMSE	Time Taken
GradientBoostingRegressor	0.90	0.90	4262.66	0.19
HistGradientBoostingRegressor	0.89	0.89	4513.08	0.63
LGBMRegressor	0.88	0.89	4548.47	0.10
BaggingRegressor	0.88	0.88	4620.42	0.07
RandomForestRegressor	0.88	0.88	4664.23	0.50
XGBRegressor	0.87	0.87	4901.26	0.15
KNeighborsRegressor	0.86	0.86	4989.34	0.02
ExtraTreesRegressor	0.85	0.85	5273.32	0.40
PoissonRegressor	0.83	0.83	5511.73	0.02
AdaBoostRegressor	0.83	0.83	5570.55	0.06
Lars	0.80	0.81	5957.61	0.01
TransformedTargetRegressor	0.80	0.81	5957.61	0.02
LinearRegression	0.80	0.81	5957.61	0.01
Lasso	0.80	0.81	5958.16	0.01

GradientBoostingRegressor

```
GradientBoostingRegressor(learning_rate=0.01, loss='absolute_error',
                          min_samples_split=5, n_estimators=1000)
```

```
rmse= np.sqrt(mean_squared_error(y_test, x_test_std))

Model Accuracy: 0.822
The mean squared error (MSE) on test set: 32671538.8388
```

	predictor	importancia
0	age	0.45
4	smoker	0.27
3	children	0.08
5	region	0.07
2	bmi	0.06
1	sex	0.06

- Our data includes mostly **overweight, middle-aged clients with healthy smoking habits and no children**.
- The trend lines of **annual insurance premiums are mostly conditioned by smoking and age** (remembering that an **unhealthy BMI is always the basis in most cases**, and is an important driver of insurance premiums).

Of these, **smoking (**difficult to monitor** and has **effects** in health for a **long time**) and **BMI** (**easily monitored** and you can see **healthy beneficts in short time**) are variables, **age**, **impossible** to change, **children**, **innapropiate**. **

- **Our model**, is able to predict with **accuracy: 0.826** and mean squared error (**MSE**) on test set: **31916978** the annual insurance premium with these variables.

PROPOSAL

1.Promote interaction of customers and prospective customers by calculating their insurance premiums and their decrease by improving their BMI
2.Facilitating access to slimming programmes for clients and future clients
3.Check annual monitoring of clients' maintenance of new BMI and healthy habits
4.Promoting healthy smoking habits (further benefits)

<https://github.com/AnaGonBu>

NEXT STEPS

1.Develop a model to predict future spending after a year of declining BMI
2.Calculate benefit for the year during the insured's BMI decline while maintaining the old BMI premium.
3.Determine time in which health expenditure actually decreases with decreasing BMI