SUPSI

US Medical charges analysis

| Studente/i	Relatore	Correlatore
Gregorio Piqué	Cannelli Loris	

Corso di laurea	Modulo / Codice Progetto	Anno
Computer Science	Machine Learning	2022/2023

Committente	Data	
	May 5, 2023	

STUDENTSUPSI

Dataset

- US citizens

- One person per observation

- **1338 rows × 7 columns**

- Data of various types:
    - demographic/geographic
    - physical
   - family
    - behavioral

|	age	sex	bmi	children	smoker	region	charges
0	19	female	27.900	0	yes	southwest	16884.92400
1	18	male	33.770	1	no	southeast	1725.55230
2	28	male	33.000	3	no	southeast	4449.46200
3	33	male	22.705	0	no	northwest	21984.47061
4	32	male	28.880	0	no	northwest	3866.85520
...	...	...	...	...	...	...	...
1333	50	male	30.970	3	no	northwest	10600.54830
1334	18	female	31.920	0	no	northeast	2205.98080
1335	18	female	36.850	0	no	southeast	1629.83350
1336	21	female	25.800	0	no	southwest	2007.94500
1337	61	female	29.070	0	yes	northwest	29141.36030

Distribution by age and sex
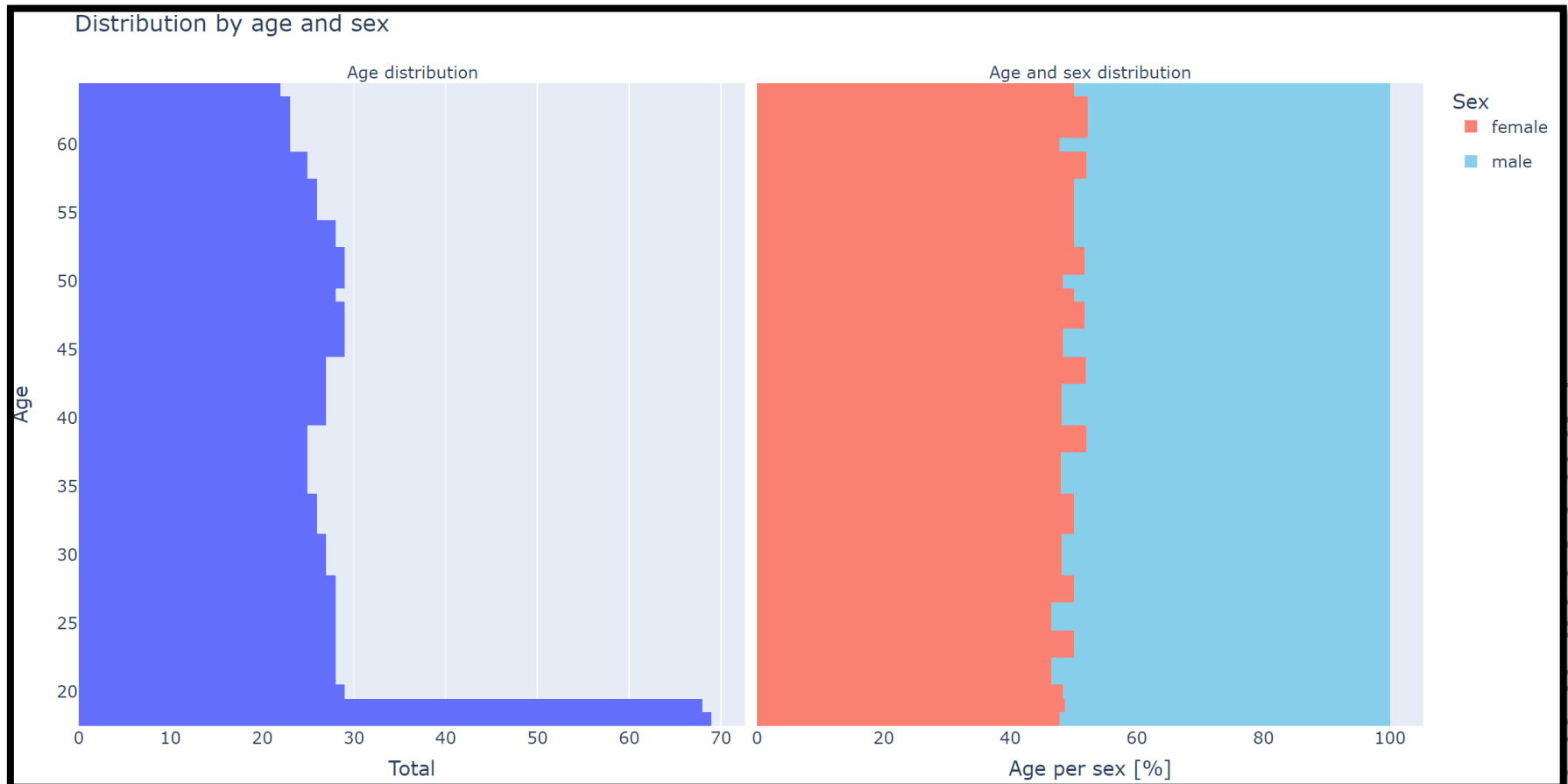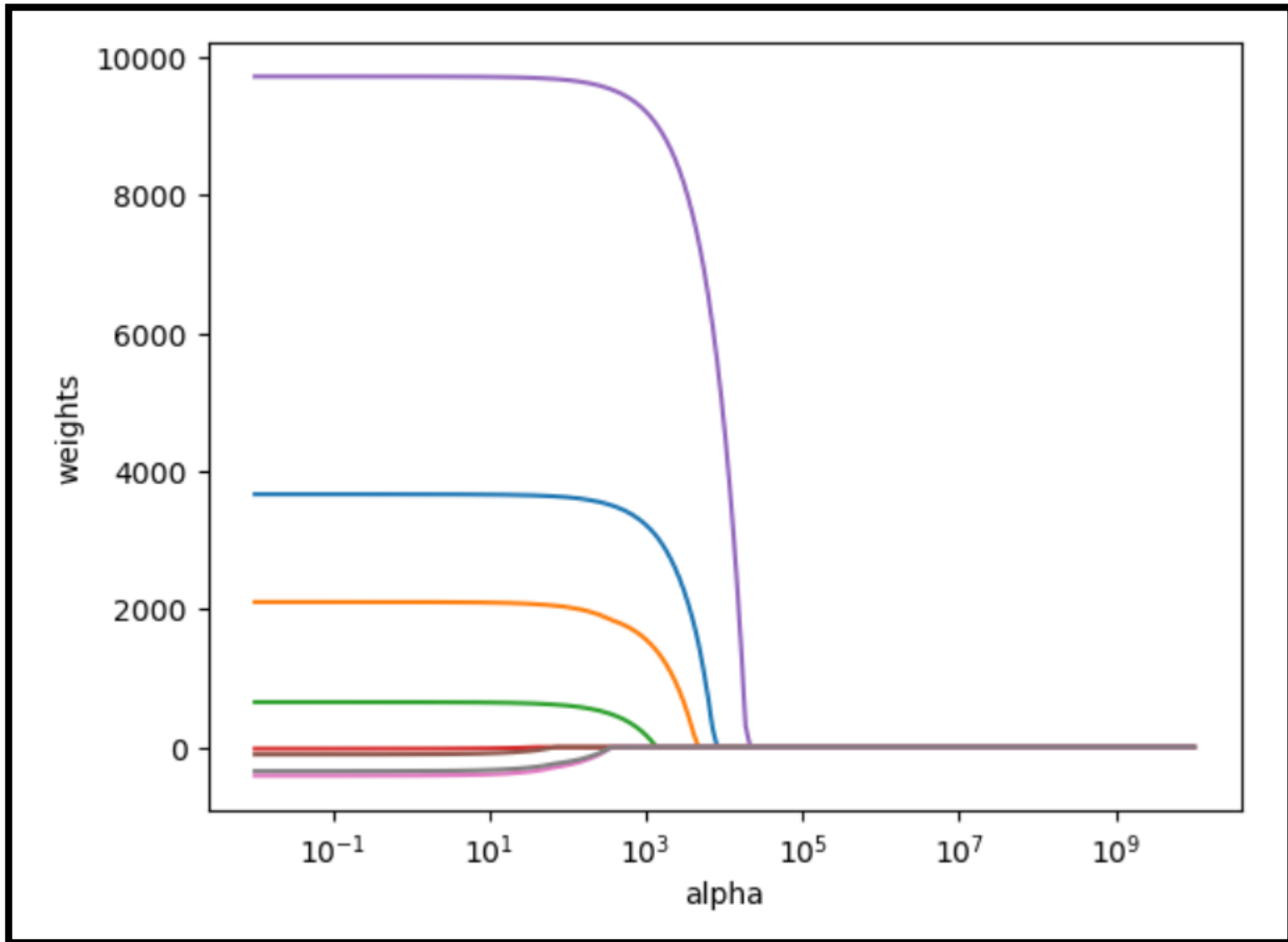
Medical Charges Prediction

Correlation
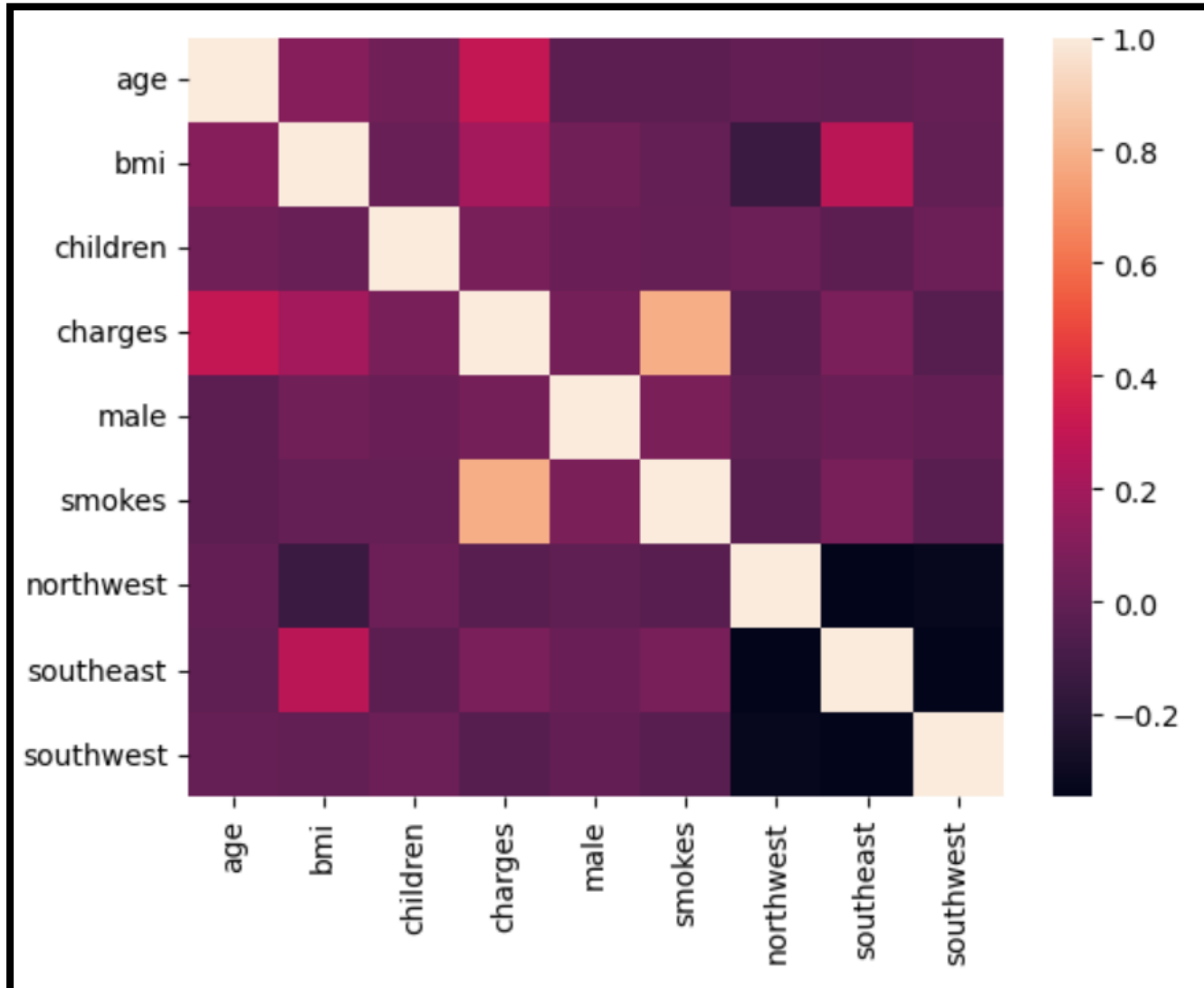
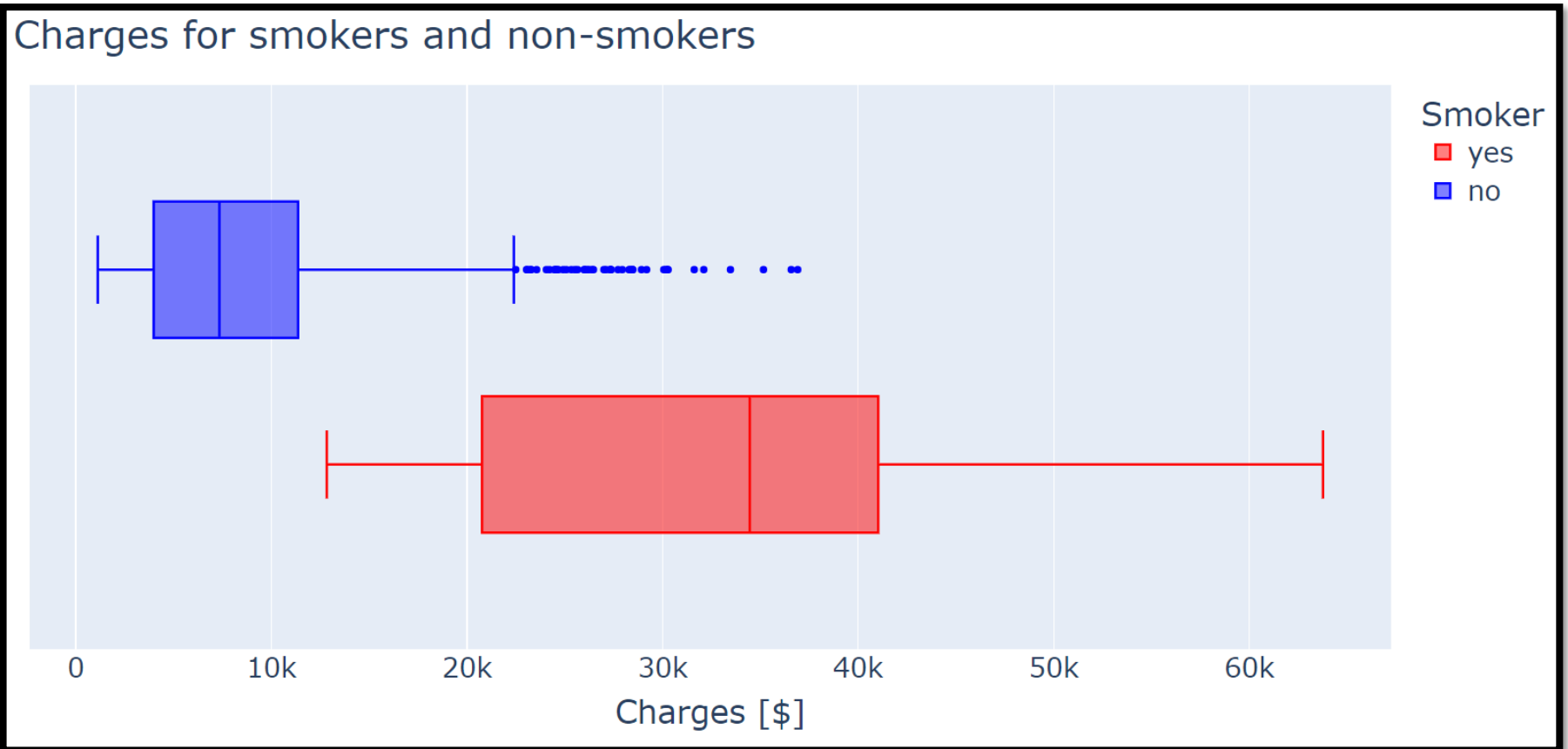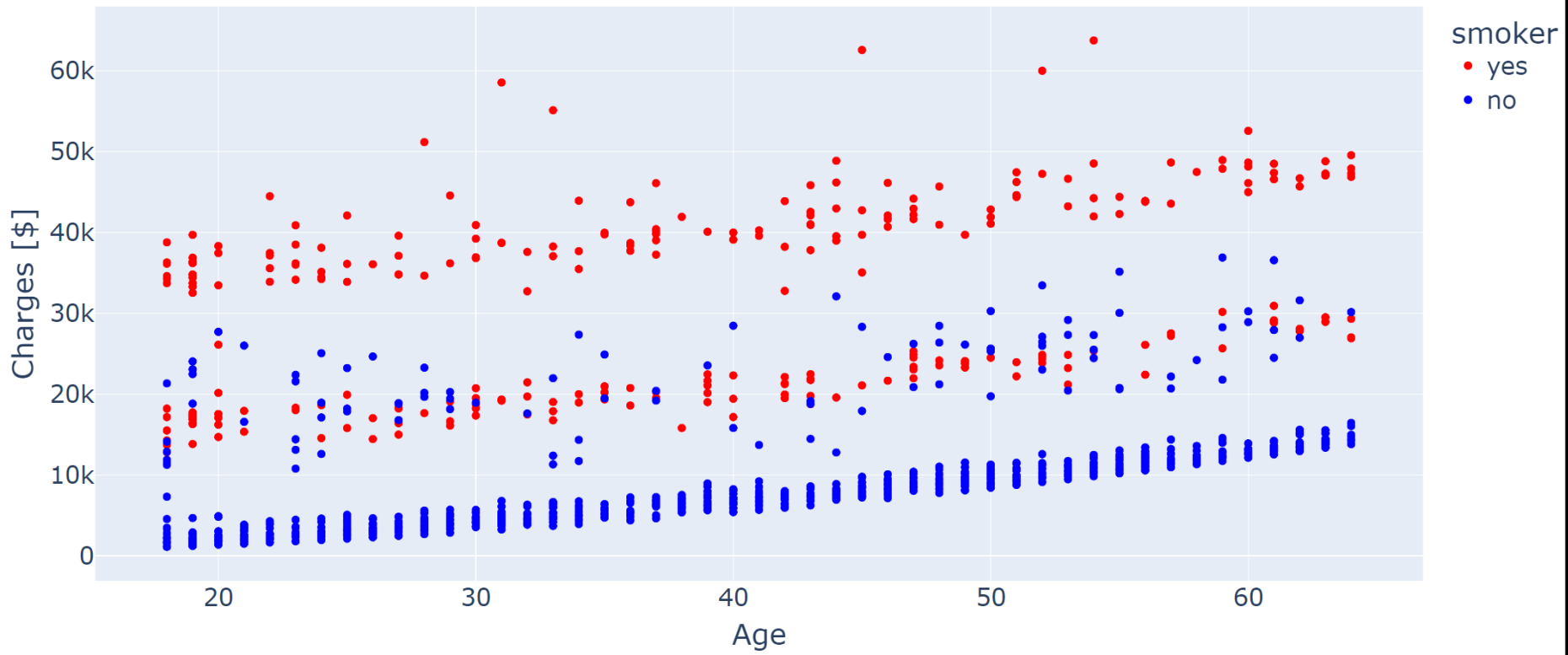# Correlation
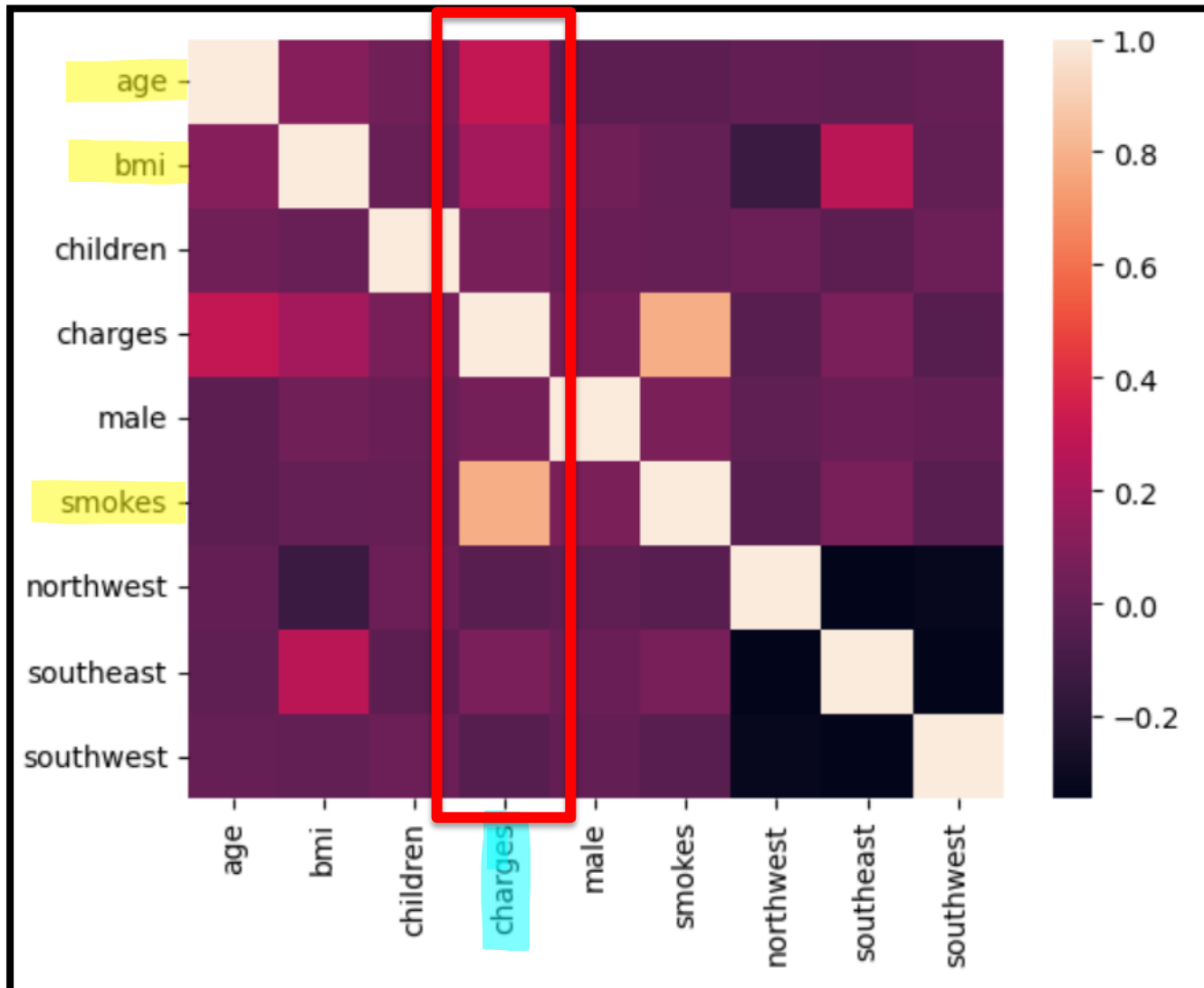
Lasso for attribute condensing

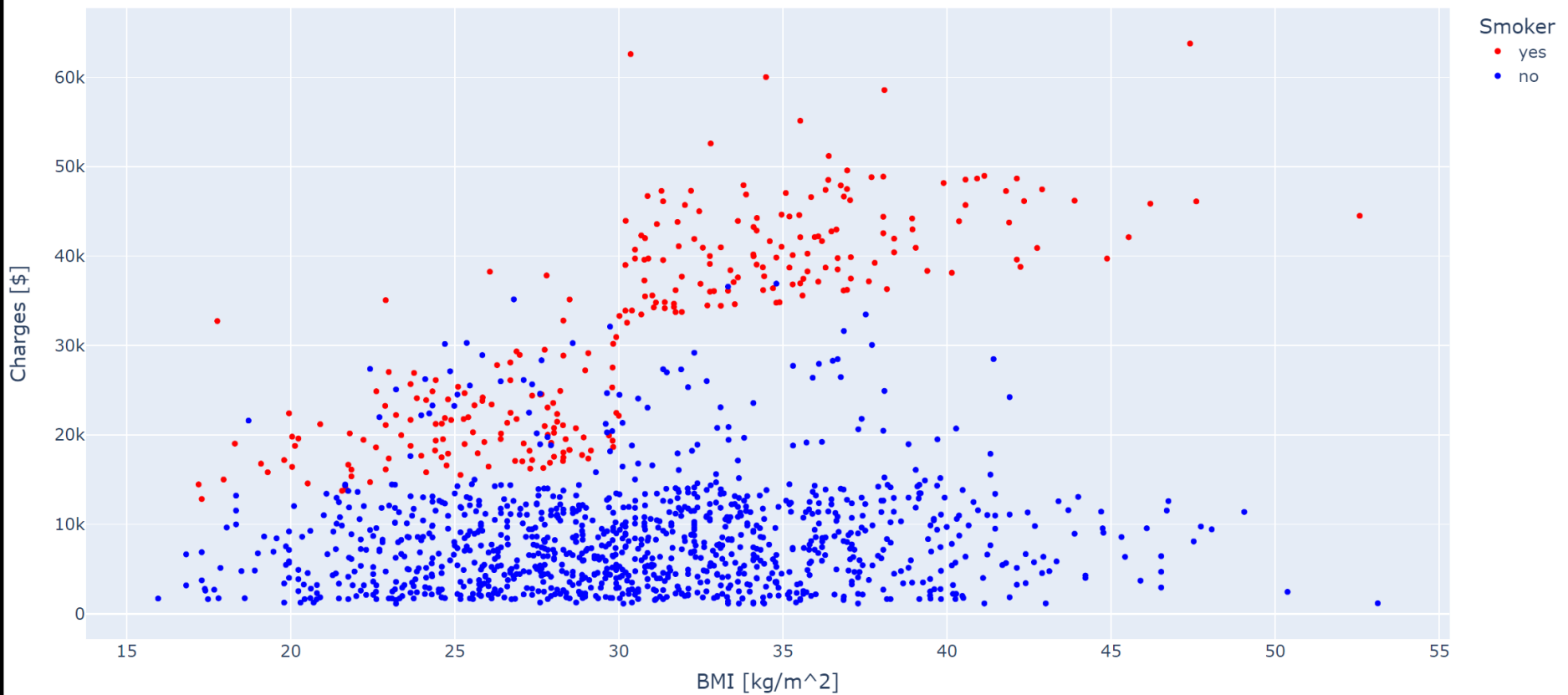# Charges for smokers and non-smokers

Charges for smokers and non-smokers
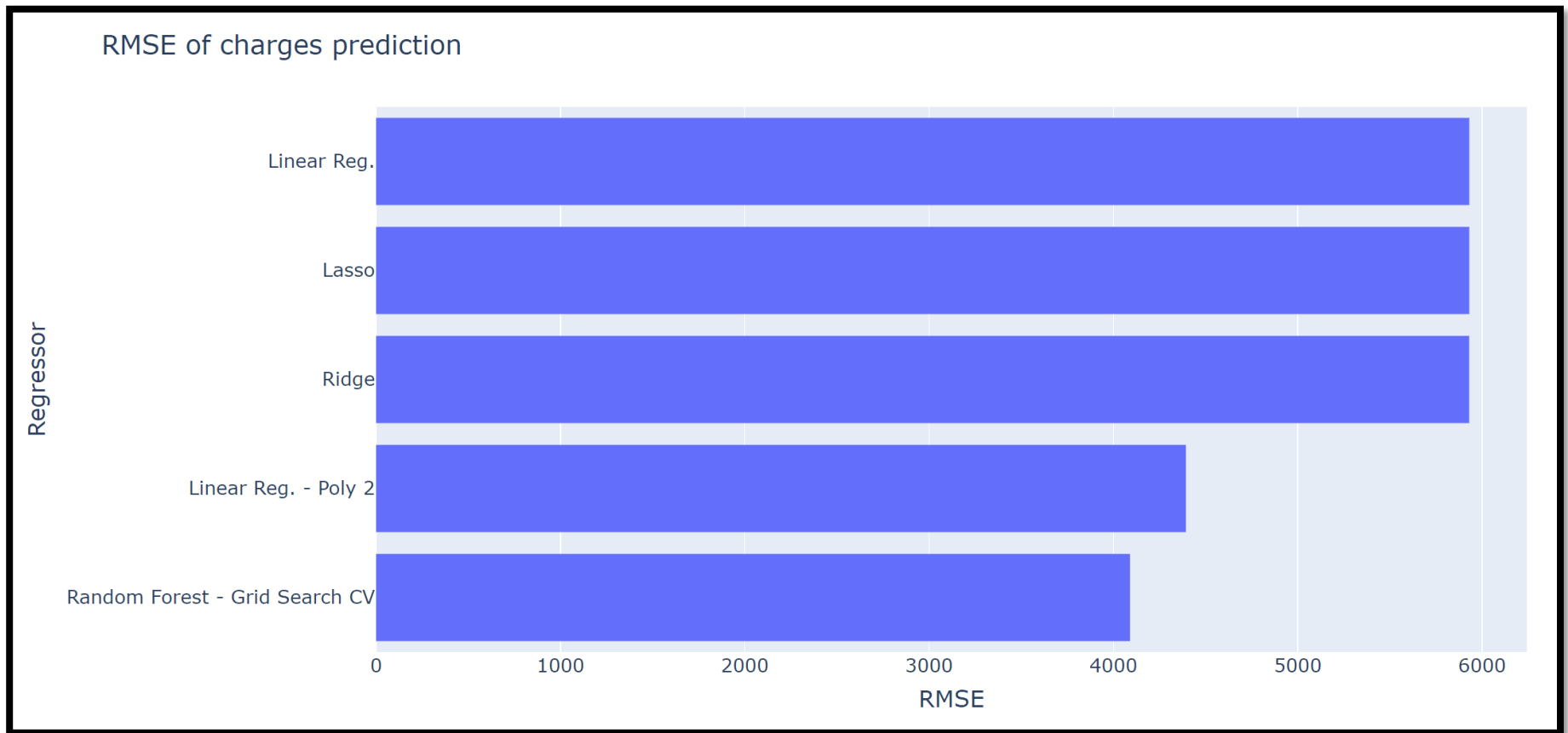
Charges by age for smokers and non-smokers
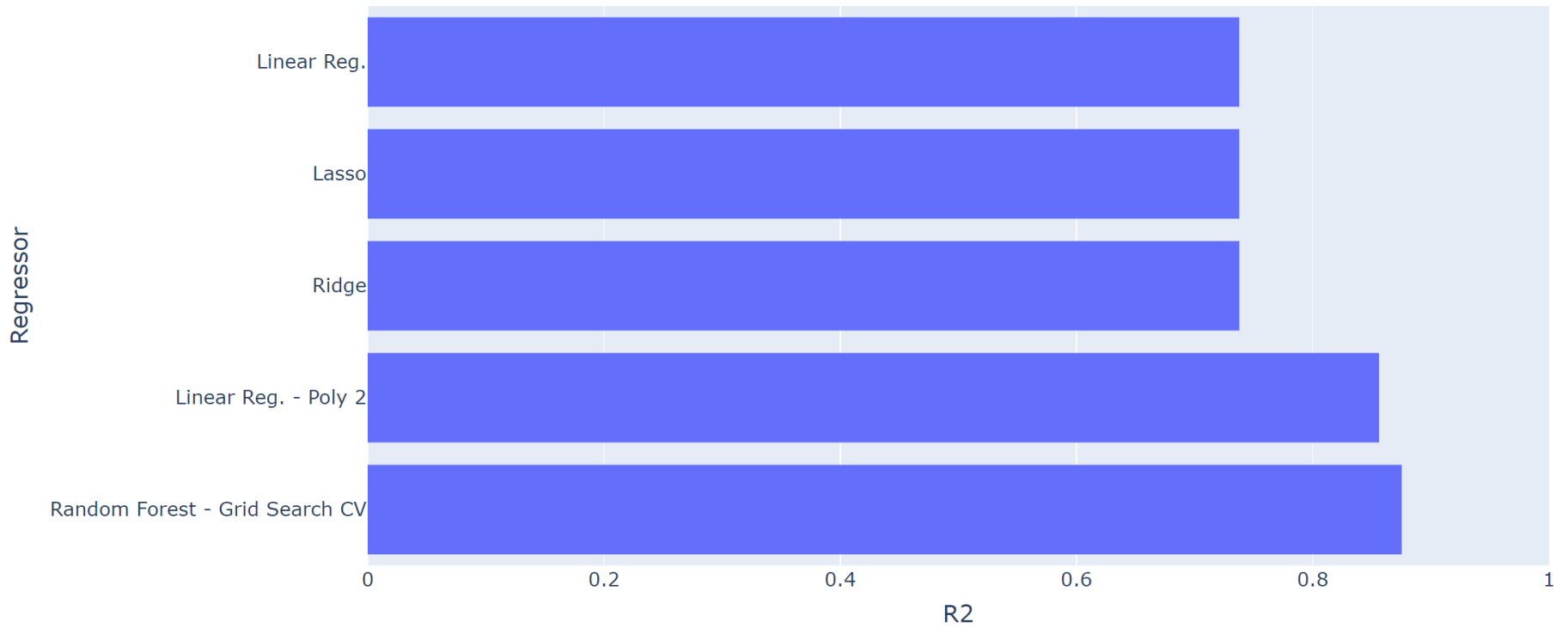
Charges by BMI for smokers and non-smokers
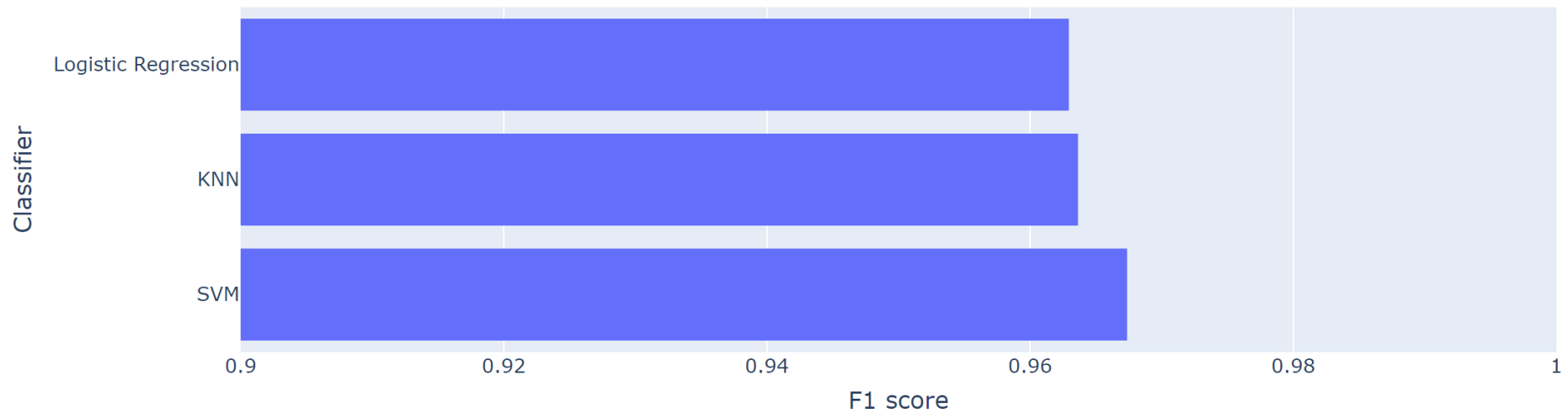
# Regression results

RMSE

RMSE of charges prediction

R2

R2 of charges prediction

Classification: Smokers

Result of prediction if person smokes

Bar chart — F1 score by Classifier:
- Logistic Regression: ≈0.963
- KNN: ≈0.963
- SVM: ≈0.968

X-axis: F1 score (0.9 to 1)
Y-axis: Classifier

	precision	recall	f1-score	support
no	1.00	0.96	0.98	215
yes	0.85	1.00	0.92	53
accuracy			0.97	268
macro avg	0.93	0.98	0.95	268
weighted avg	0.97	0.97	0.97	268

Classification: Sex

Charges by BMI for sex
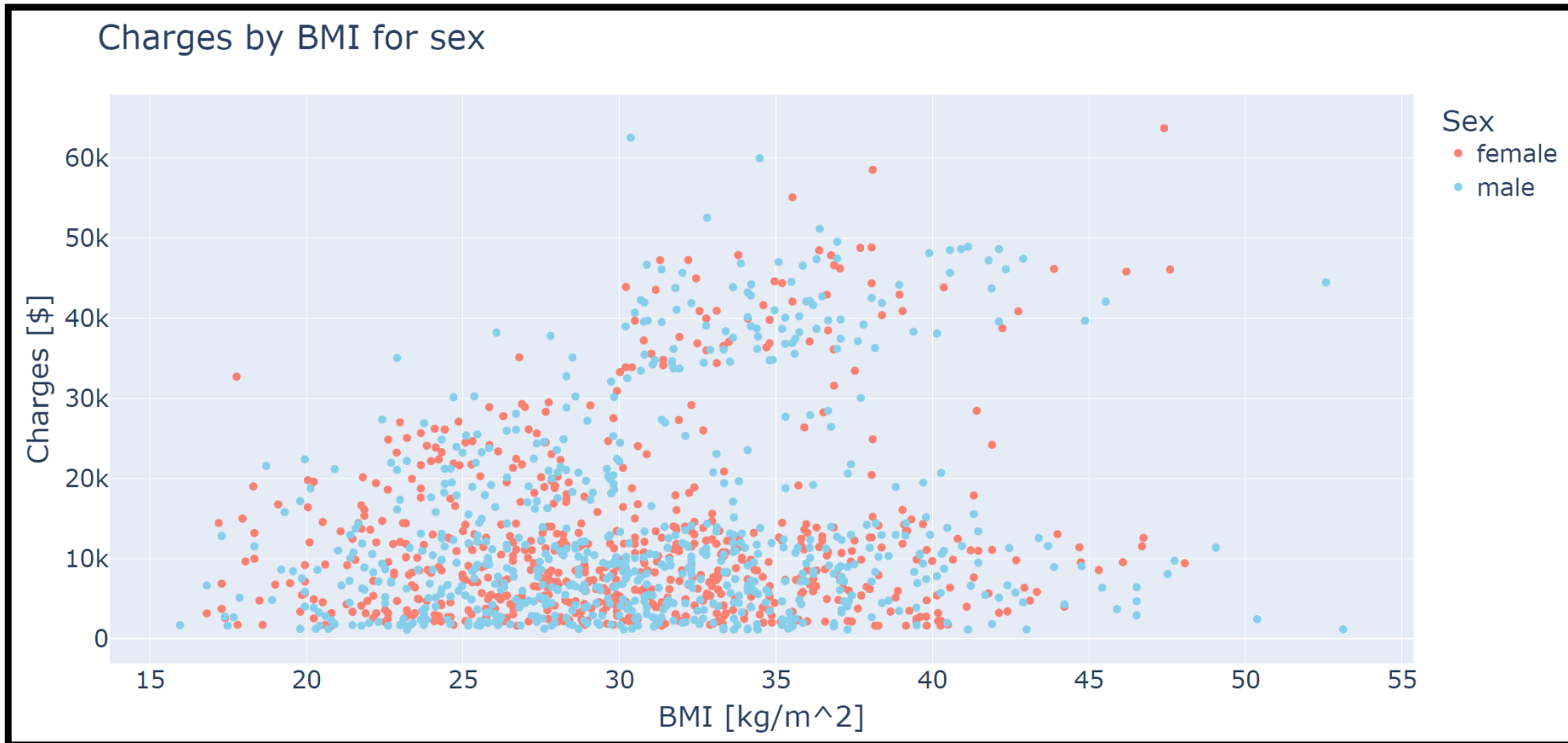
Result of prediction of person's sex
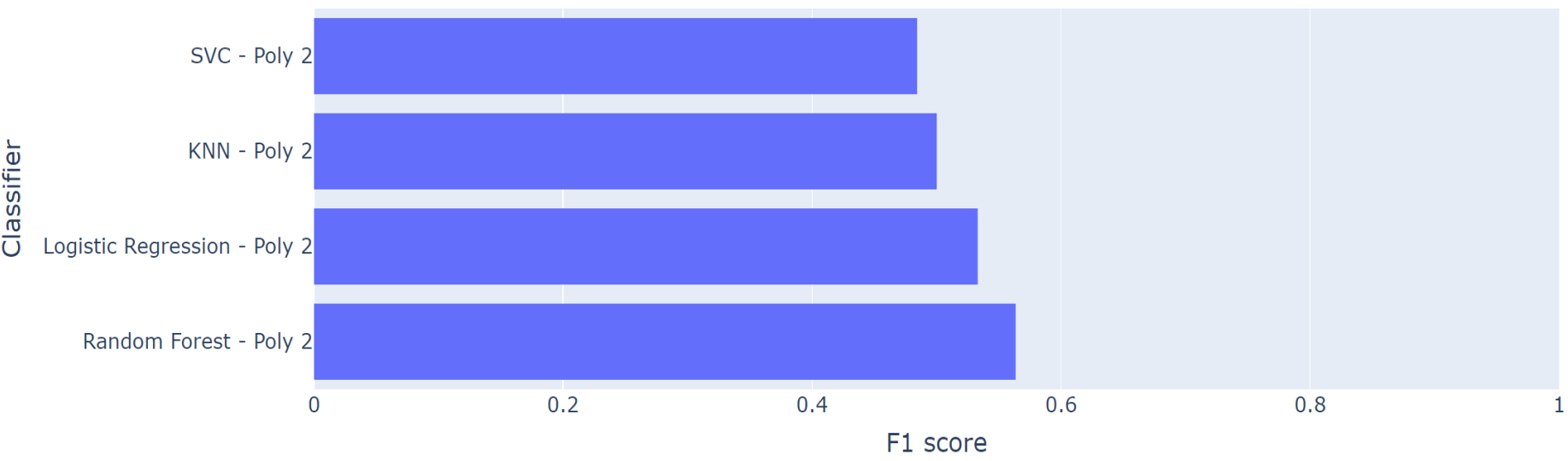
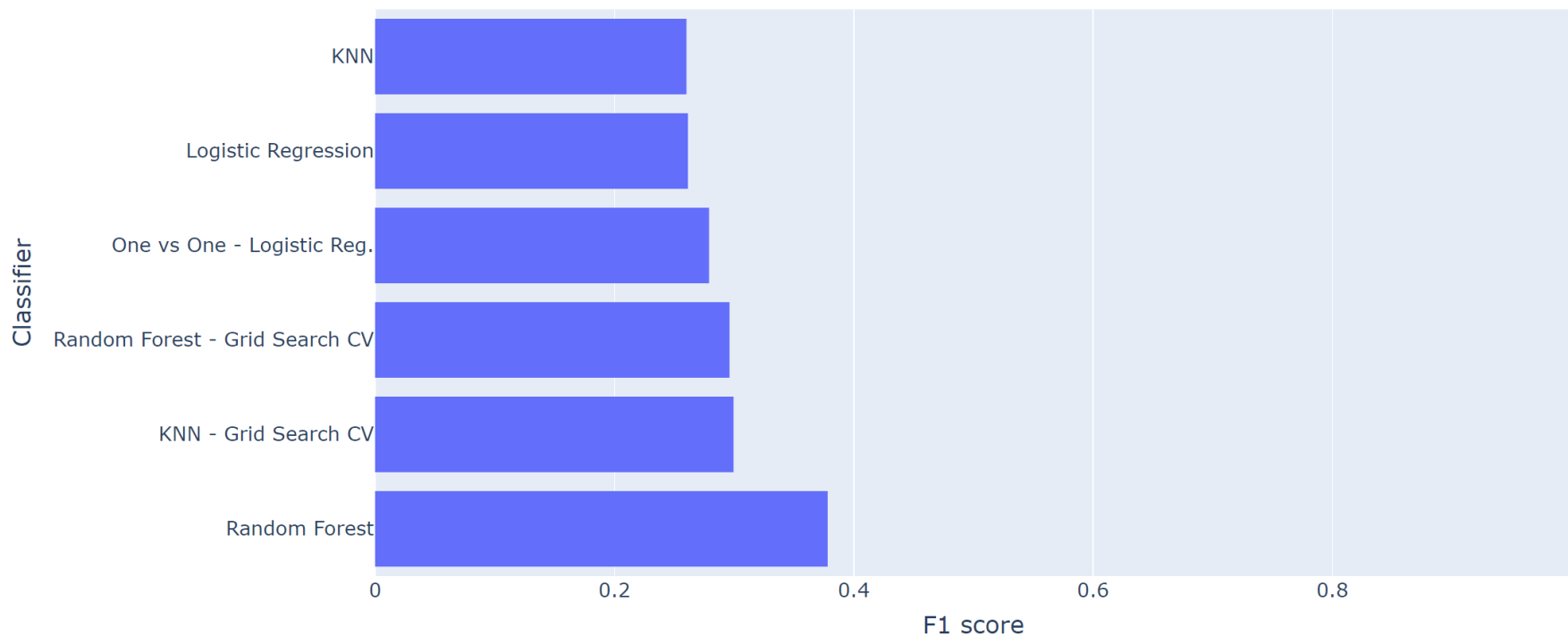|	precision	recall	f1-score	support
male	0.55	0.58	0.56	130
female	0.58	0.55	0.57	138
accuracy			0.56	268
macro avg	0.56	0.56	0.56	268
weighted avg	0.56	0.56	0.56	268

Classification: Region

Charges by region (and national)

Charges by region (and national)
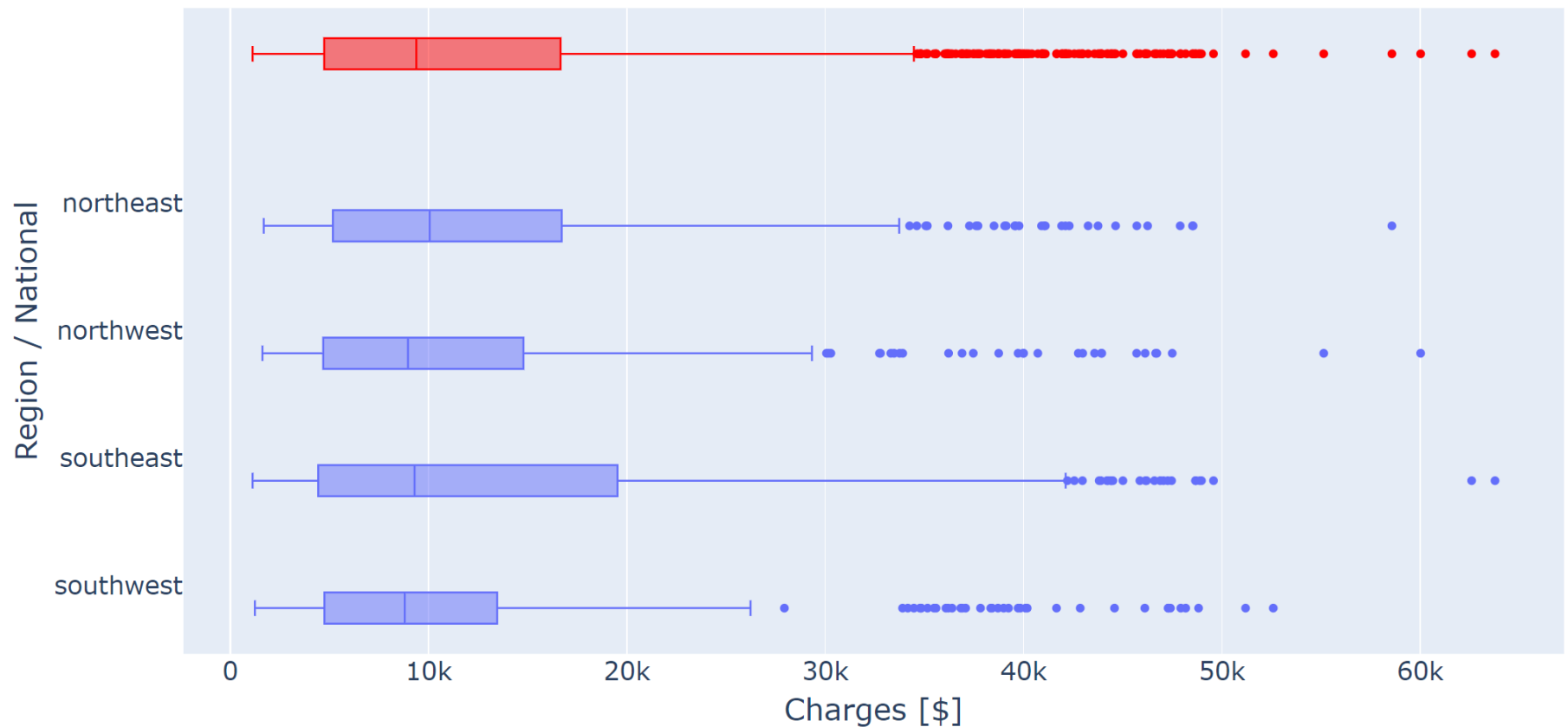
Result of prediction of person's region

	precision	recall	f1-score	support
northwest	0.31	0.39	0.35	56
southwest	0.34	0.32	0.33	63
northeast	0.46	0.40	0.43	80
southeast	0.38	0.39	0.39	69
accuracy			0.38	268
macro avg	0.38	0.38	0.37	268
weighted avg	0.38	0.38	0.38	268