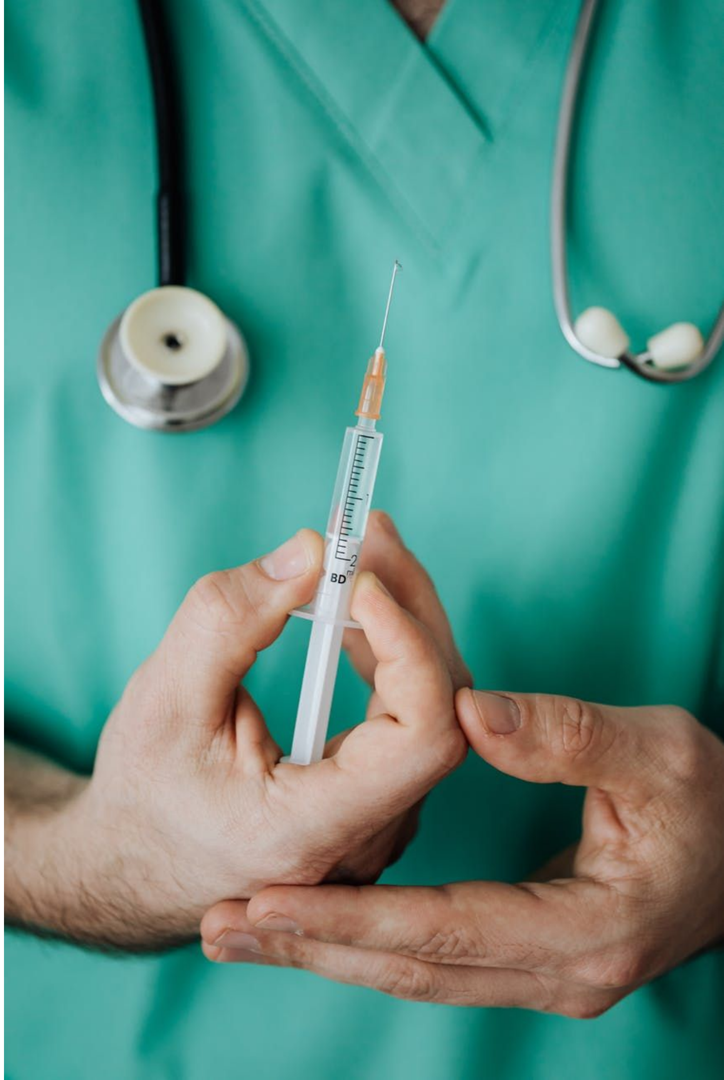
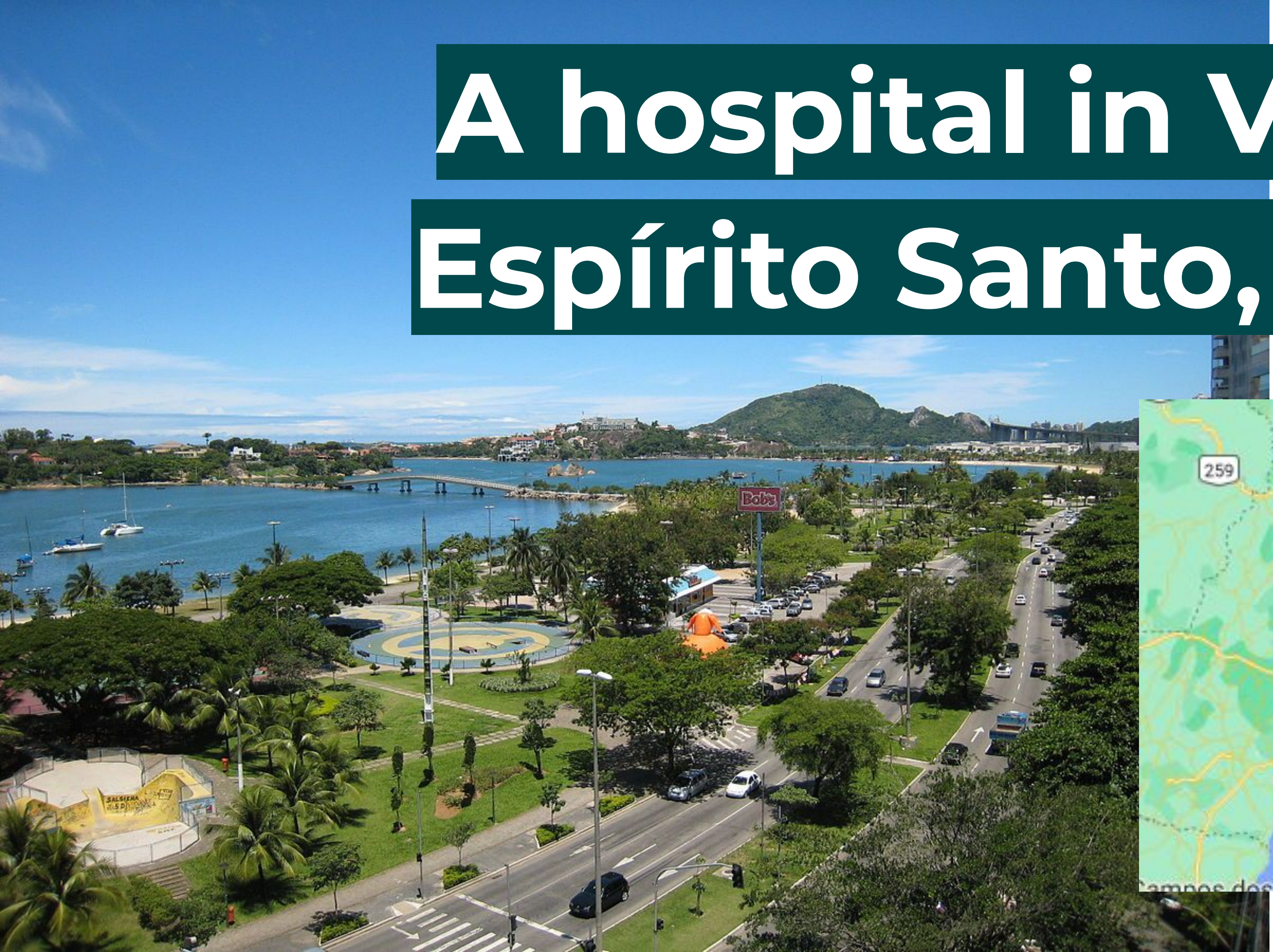
MEDICAL APPOINTMENT NO SHOW PREDICTION

Final Project of Data Science Course

Gloria Aprilia (JCDS-10)

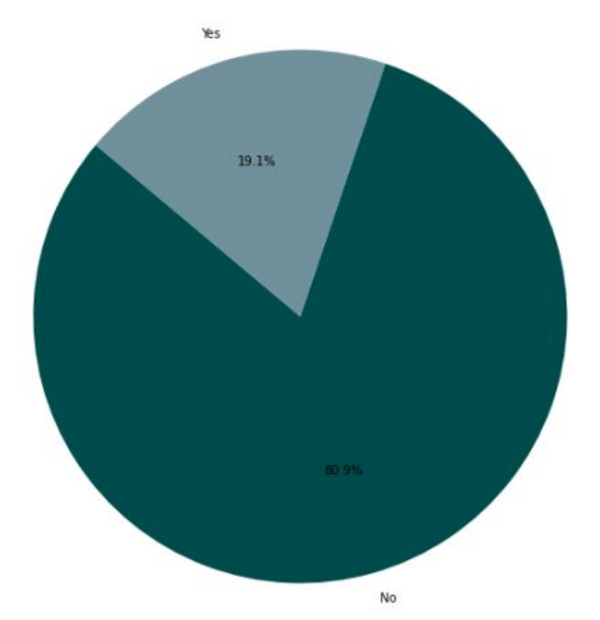


A hospital in Vitória, Espírito Santo, Brazil



Problem:

No show case reaches nearly 20% of total appointment



Impact:

1. Hospital's revenue loss, approximately BRL6,177,120 (USD1,153,718) a month
2. Doctor salary loss
3. Patient treatment delay, that may lead to more severe illness

Goals

1. **Find the factors** of outpatient no show behavior
2. **Find quantitative measure** that represents the factors
3. **Build machine learning model** to predict no show patient
4. **Connect model** with dashboard

Methods

Data Cleansing

missing value removal

redundant data removal

outliers handling

EDA

extract insights

prepare appropriate data for ML modeling

Machine Learning

Logistic regression

K-Nearest Neighbor

Random Forest

XGBoost

Dashboard

show data sample

display graphs

perform prediction

Literature Review

3 major factor of no show[1]

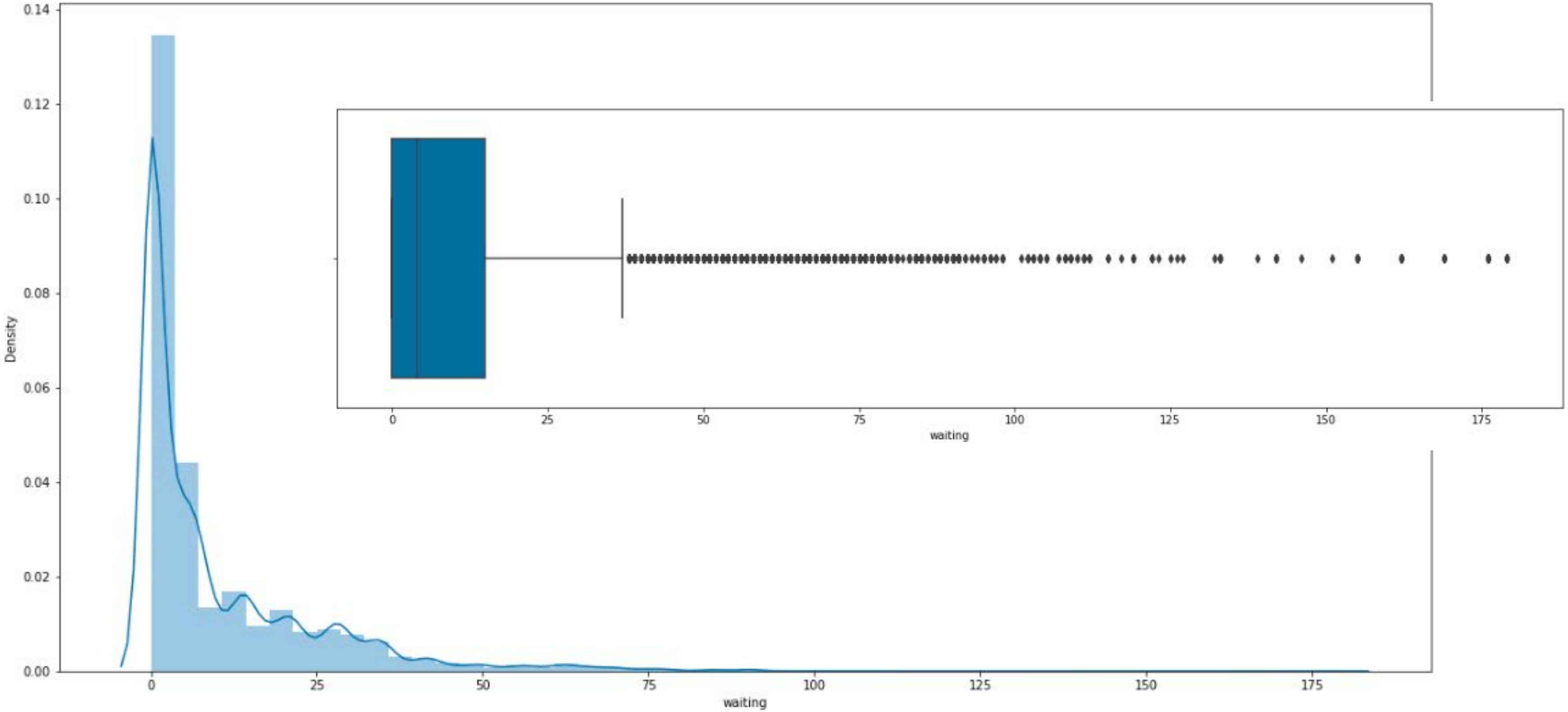
- emotional condition of patient
- perceived disrespect
- not understanding scheduling system

4 major factor of no show[2]

- patient-related issue
- environmental issue
- financial issue
- scheduling-related issue

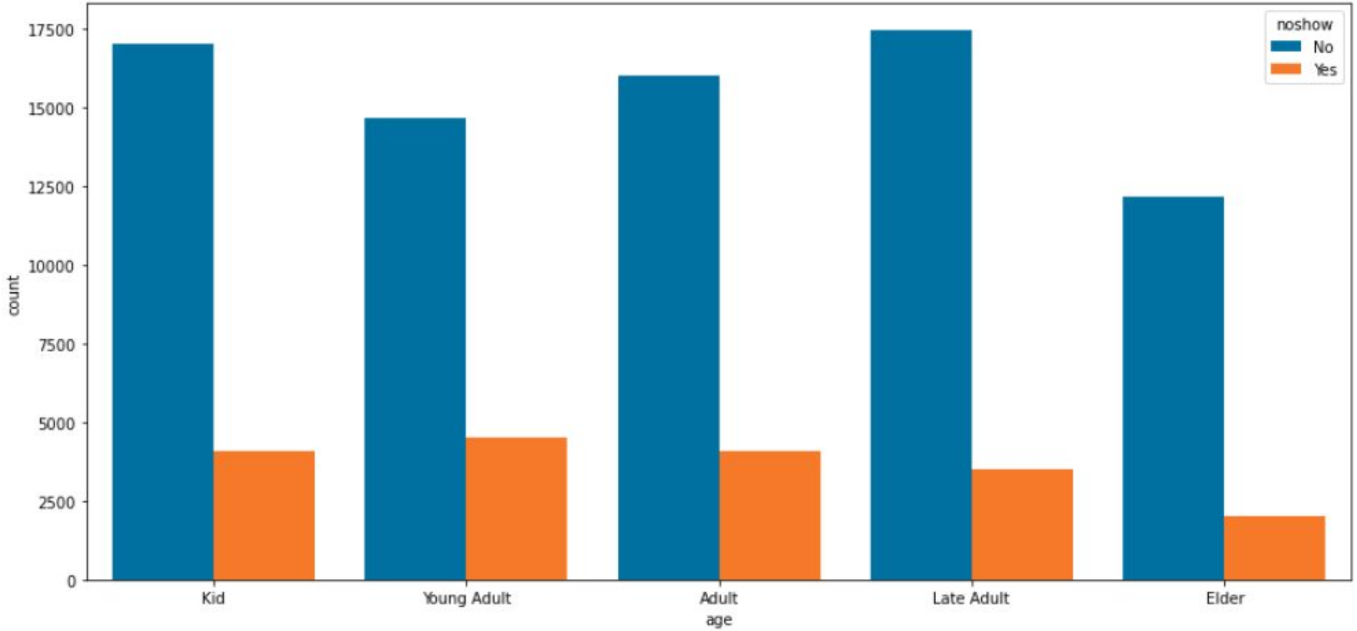
[1] Lacy, N. L., Paulman, A., Reuter, M. D., & Lovejoy, B. (2004). Why we don't come: patient perceptions on no-shows. The Annals of Family Medicine, 2(6), 541-545.
[2] Marbouh, D., Khaleel, I., Al Shanqiti, K., Al Tamimi, M., Simsekler, M. C. E., Ellahham, S., ... & Alibazoglu, H. (2020). Evaluating the Impact of Patient No-Shows on Service Quality. Risk Management and Healthcare Policy, 13, 509-517.

Data Cleansing



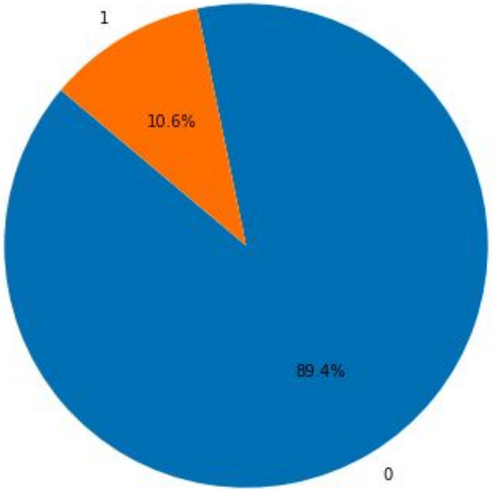
There are 90.61% of total patients scheduled appointments within 30 days

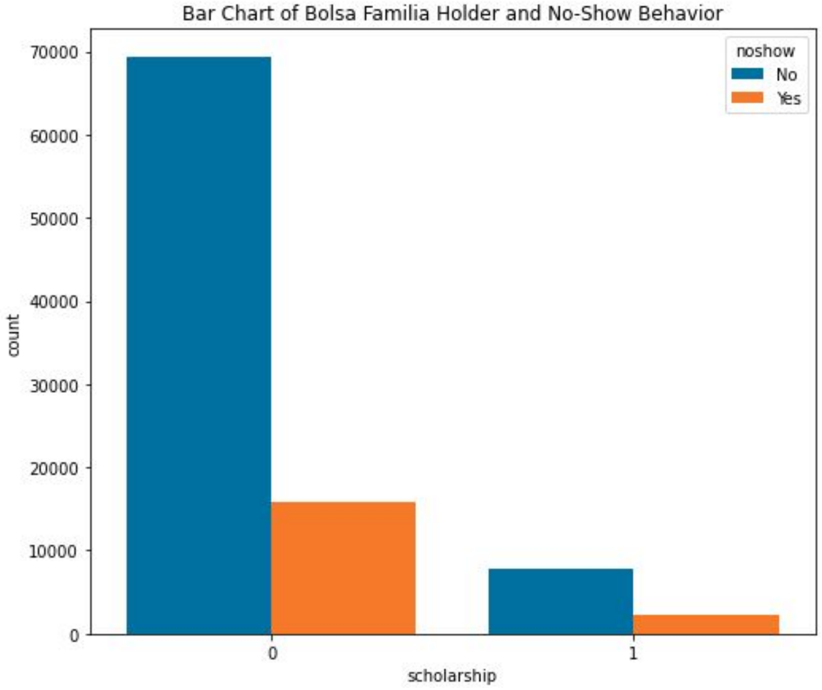
Exploratory Data Analysis



noshow / age	No	Yes
Kid	80.67	19.33
Young Adult	76.46	23.54
Adult	79.70	20.30
Late Adult	83.26	16.74
Elder	85.70	14.30

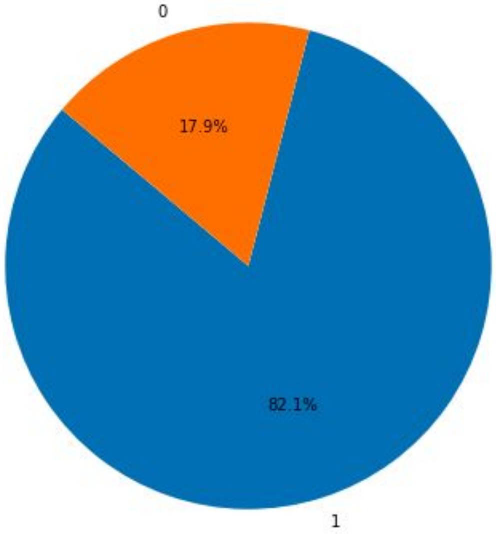
Exploratory Data Analysis

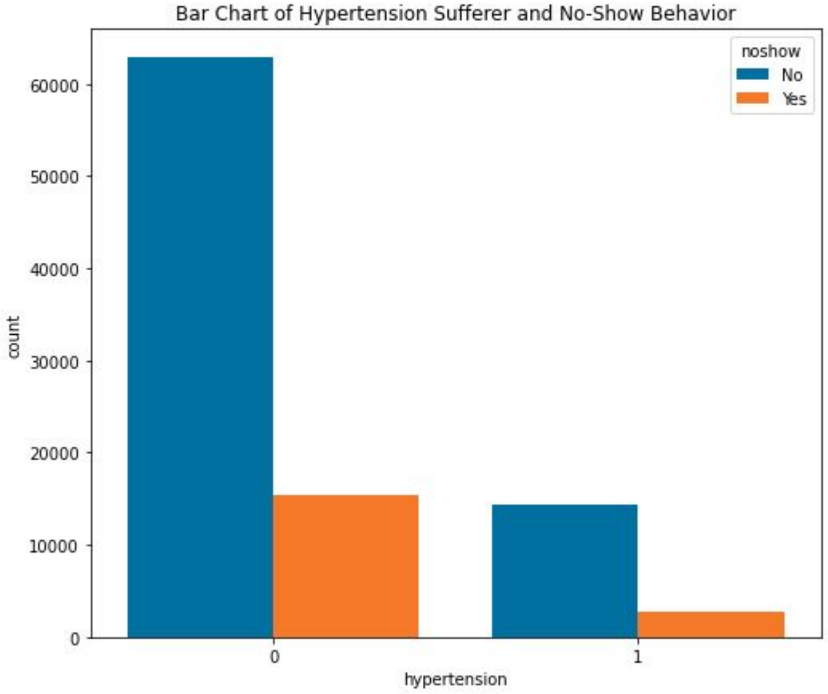




noshow	No	Yes
scholarship		
0	81.34	18.66
1	77.55	22.45

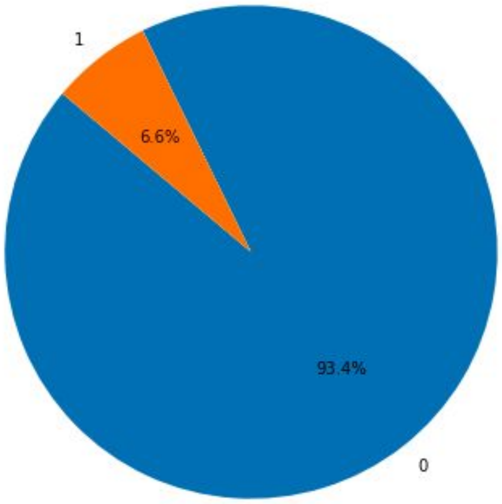
Exploratory Data Analysis

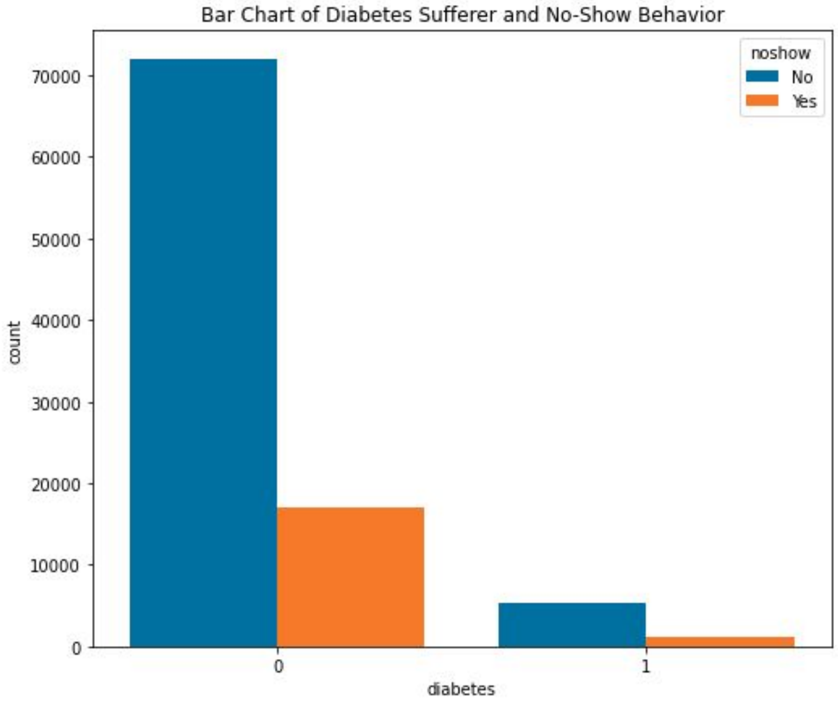




noshow	No	Yes
hypertension		
0	80.34	19.66
1	83.65	16.35

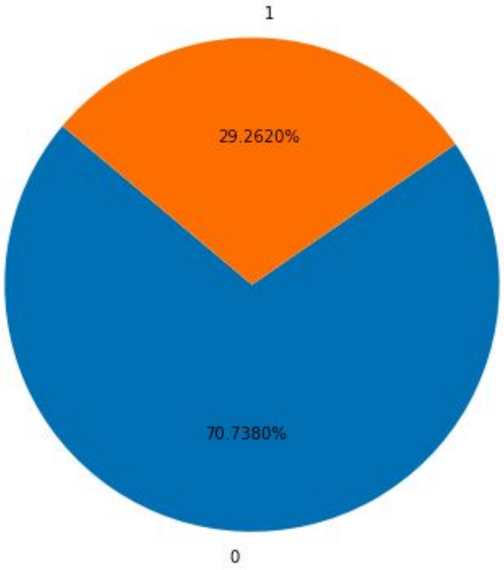
Exploratory Data Analysis

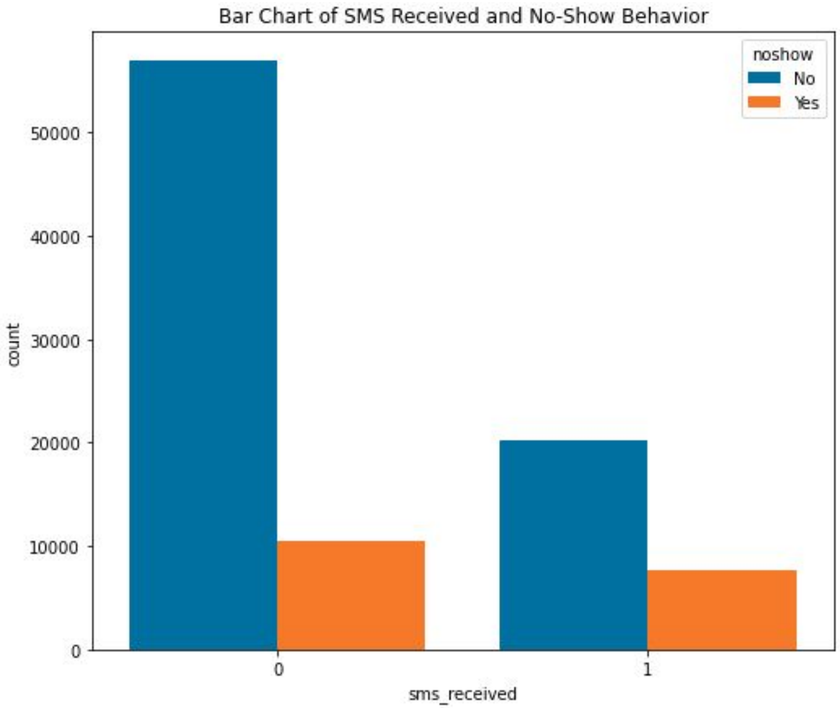




noshow	No	Yes
diabetes		
0	80.81	19.19
1	82.68	17.32

Exploratory Data Analysis



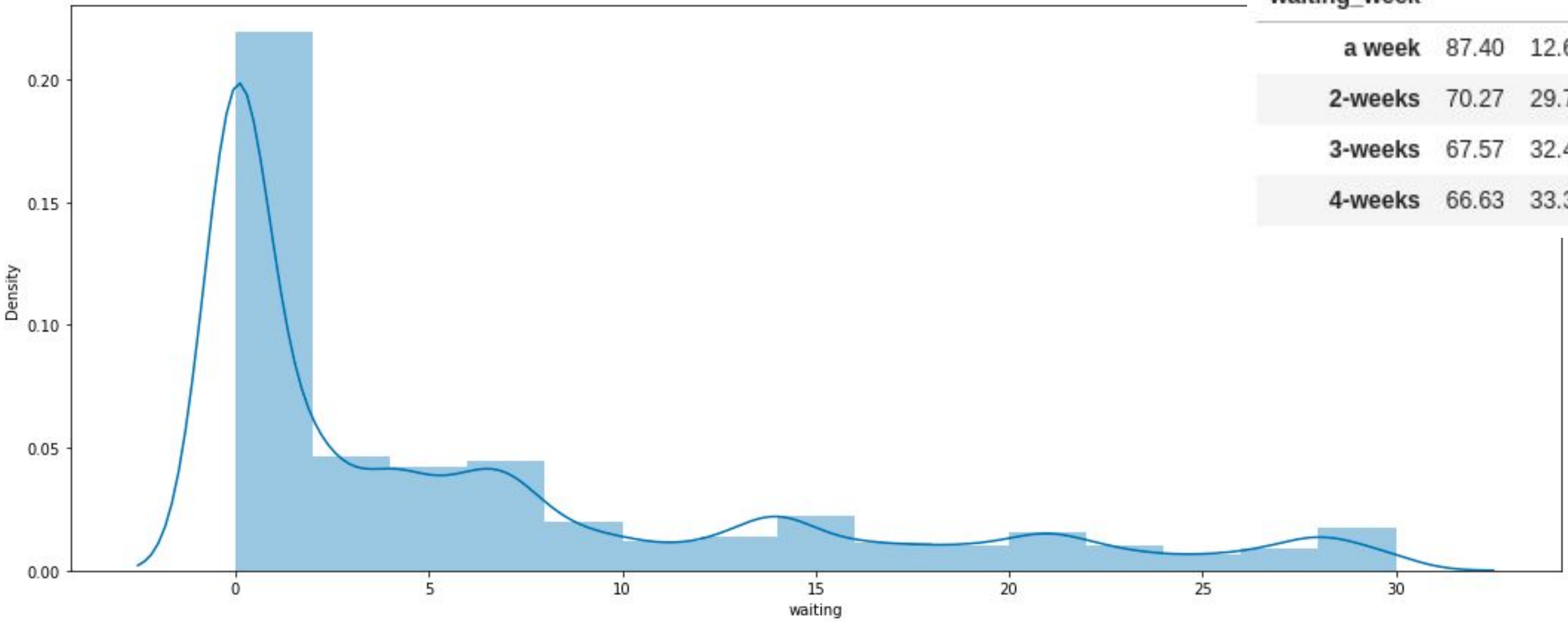


noshow	No	Yes
sms_received		
0	84.35	15.65
1	72.69	27.31

	noshow	No	Yes
sms_received	**waiting_week**		
0	**a week**	89.04	10.96
	2-weeks	67.57	32.43
	3-weeks	63.98	36.02
	4-weeks	61.14	38.86
1	**a week**	76.64	23.36
	2-weeks	72.18	27.82
	3-weeks	70.02	29.98
	4-weeks	69.91	30.09

Exploratory Data Analysis

noshow	No	Yes
waiting_week		
a week	87.40	12.60
2-weeks	70.27	29.73
3-weeks	67.57	32.43
4-weeks	66.63	33.37



Exploratory Data Analysis

```
# see the correlation between features and noshow column
df_enc.corr()['noshow'].sort_values(ascending=False)[1:]
```

```
waiting                  0.234944
sms_received             0.134966
scholarship              0.029709
sched_weekday            0.007205
weather_partly sunny     0.007187
appt_day                 0.006144
sched_day                0.001425
alcoholism               0.000109
appt_weekday            -0.000188
neighborhood            -0.001088
weather_cloudy          -0.001214
weather_sunny           -0.003628
handicap                -0.005735
gender                  -0.007588
diabetes                -0.011860
hypertension            -0.032342
age                     -0.050129
Name: noshow, dtype: float64
```

	waiting	sms_received	scholarship	diabetes	hypertension	age	noshow
0	0	0	0	0	1	62	0
1	0	0	0	0	0	56	0
2	0	0	0	0	0	62	0
3	0	0	0	0	0	8	0
4	0	0	0	1	1	56	0

Machine Learning Modeling

Highlight:

1. Age is scaled using MinMax scaler
2. Waiting time is not scaled to give it more weight
3. Data balancing using SMOTE
4. Test size: 19,061 data
5. Scoring priority: recall and ROC

Result

	accuracy	precision	recall	f1	ROC
Logistic Regression	0.686585	0.318379	0.564392	0.407106	0.691502
Logistic Reg. Tuned	0.686585	0.318379	0.564392	0.407106	0.691502
KNN	0.715545	0.305144	0.385250	0.340550	0.661841
KNN Tuned	0.685798	0.311268	0.534397	0.393396	0.707826
Random Forest	0.691989	0.295259	0.443864	0.354622	0.646244
Random Forest Tuned	0.545092	0.280981	0.889103	0.427014	0.726649
XGB	0.685588	0.311912	0.538250	0.394952	0.709255
XGB Tuned	0.595037	0.295689	0.813429	0.433717	0.730141

Consideration

- Random Forest Tuned with high recall
- XGB Tuned with high performance

Financial Projection

ASSUMPTIONS

- Medical appointment rate is BRL340
- Handling predicted no show patient costs BRL55

WITHOUT MACHINE LEARNING

Revenue calculation per register:

$$\frac{340*(\#patient\text{-}appear)}{total\ register}$$

USING MACHINE LEARNING

Machine learning model cost:

(340*FN) + 55*(TP+FP)

Revenue calculation per register:

$$\frac{340*(\#patient\text{-}appear) - machine\ learning\ cost}{total\ register}$$

act/pred	Appear	No show
Appear	TN	FP
No show	FN	TP

Financial Projection

USING TUNED RANDOM FOREST CLASSIFIER

	Pred 1	Pred 0
Act 1	3231	403
Act 0	8268	7159

Predicted Income	6,480,740
ML Cost	-769,465
Predicted Revenue	5,711,275
Without ML Revenue	5,245,180

Saving = **BRL 466,095**

USING TUNED XGBOOST CLASSIFIER

	Pred 1	Pred 0
Act 1	2956	678
Act 0	7041	8386

Predicted Income	6,480,740
ML Cost	-780,355
Predicted Revenue	5,700,385
Without ML Revenue	5,245,180

Saving = **BRL 455,205**

Conclusion

1. **Quantitative measures** that are strongly related to no show patients are: age, waiting time, scholarship, hypertension, diabetes, and sms sent to them.
2. By predicting using **Tuned Random Forest Classifier** (recall: 89%, ROC: .723), we can save hospital revenue up to BRL 24.5 per register.

Recommendation

1. Tuned Random Forest Classifier is recommended to be applied on medical appointment registration system.
2. Recommended actions to handle no show suspect:
   - ➔ Reminder (through SMS and call)
   - ➔ Shorten the waiting time
   - ➔ Enhance patient's understanding about scheduling system and health care procedure they may experience

THANK YOU

Note:

Not all no show predicted patients can keep their appointment after we treat them with recommended actions. But at least, we have done our best.