

ANÁLISE EXPLORATÓRIA DE DADOS DE ACIDENTES AERONÁUTICOS NO BRASIL

Ana Carolina

<https://github.com/ana181084>

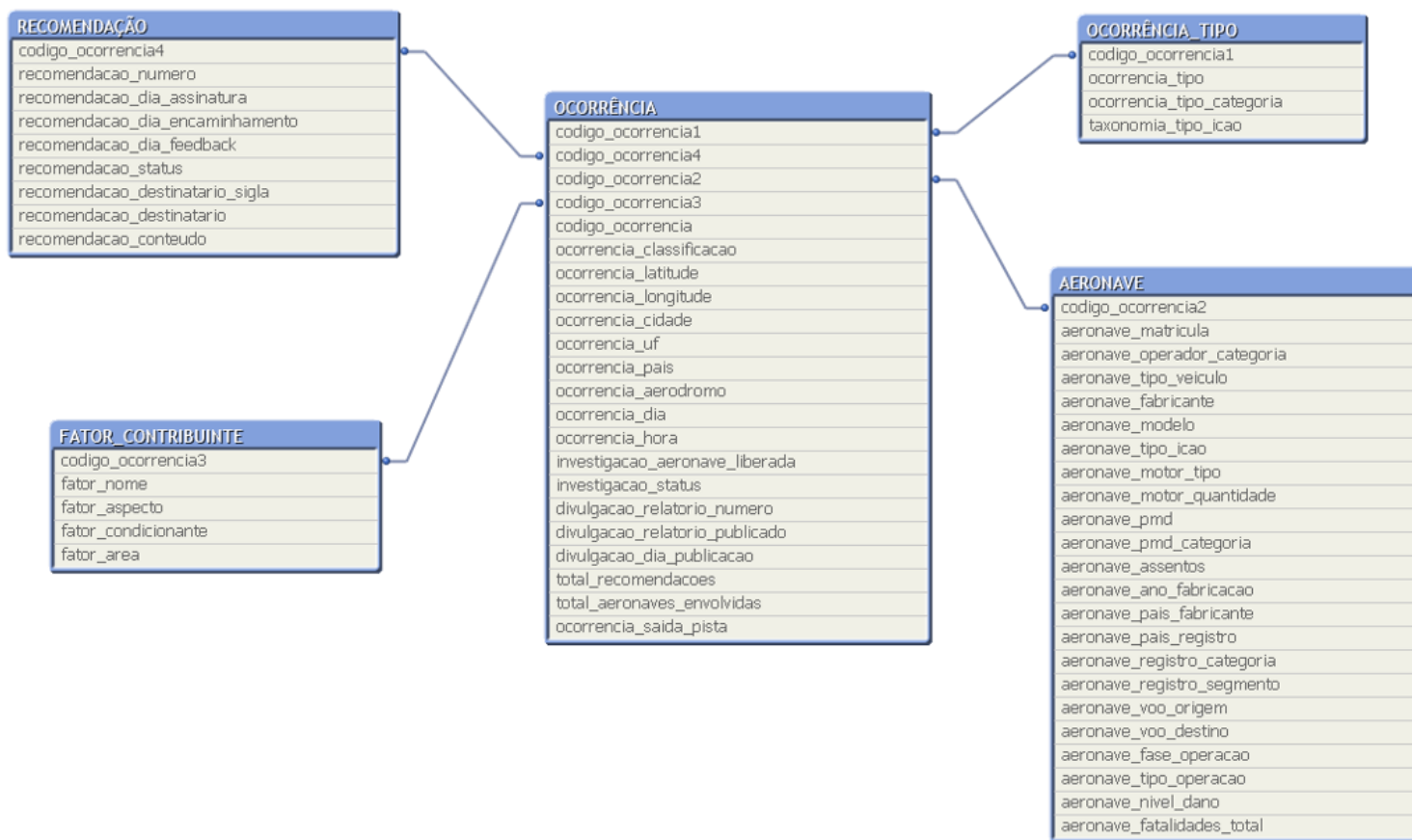
APRESENTAÇÃO DO DESAFIO

- O teste técnico consiste na exploração da base dados "Ocorrências Aeronáuticas na Aviação Civil Brasileira" dos dados abertos do governo (<https://dados.gov.br/dataset/ocorrencias-aeronauticas-da-aviacao-civil-brasileira>).
- A base de dados de ocorrências aeronáuticas é gerenciada pelo Centro de Investigação e Prevenção de Acidentes Aeronáuticos (CENIPA). Constam nesta base de dados as ocorrências aeronáuticas notificadas ao CENIPA no anos de 2012 a 2021 que ocorreram em solo brasileiro.
- O objetivo deste teste é demonstrar as habilidades como Cientista de Dados sendo criativo nas análises.

EXPLICAÇÃO DO PROCESSO UTILIZADO

- Foi realizada a análise exploratória dos dados abertos disponibilizados pelo centro de investigação e prevenção de acidentes aeronáuticos (CENIPA), onde constam 6.238 ocorrências aeronáuticas notificadas entre os anos de 2012 e 2020.
- Na segunda fase foi realizada a estruturação, adequação dos dados e posterior ajuste de modelos de machine Learning utilizando a biblioteca *pycaret* . O objetivo foi prever a classificação das ocorrências em acidente, incidente grave, incidente a partir da base de variáveis fornecidas.

EXPLICAÇÃO DO PROCESSO UTILIZADO



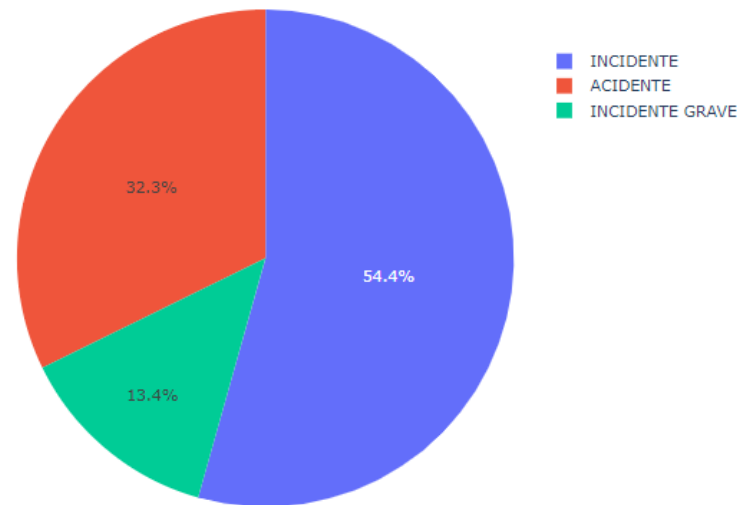
HIPOTESES LEVANTADAS

HIPOTHESES LEVANTADAS

- 1) Qual é a porcentagem de classificação das ocorrências relatadas?
- 2) Quais são os tipos de danos mais relatados nas ocorrências de acidente?
- 3) Qual são os tipos de aeronaves e o tipo de motor mais relatadas nas ocorrências de acidente?
- 4) Quais são os estados mais relatados nas ocorrências de acidente?
- 5) Quais são os tipo de ocorrências de acidente mais relatados?
- 6) Qual é o número de ocorrências de acidente por ano?

ANÁLISE EXPLORATÓRIA

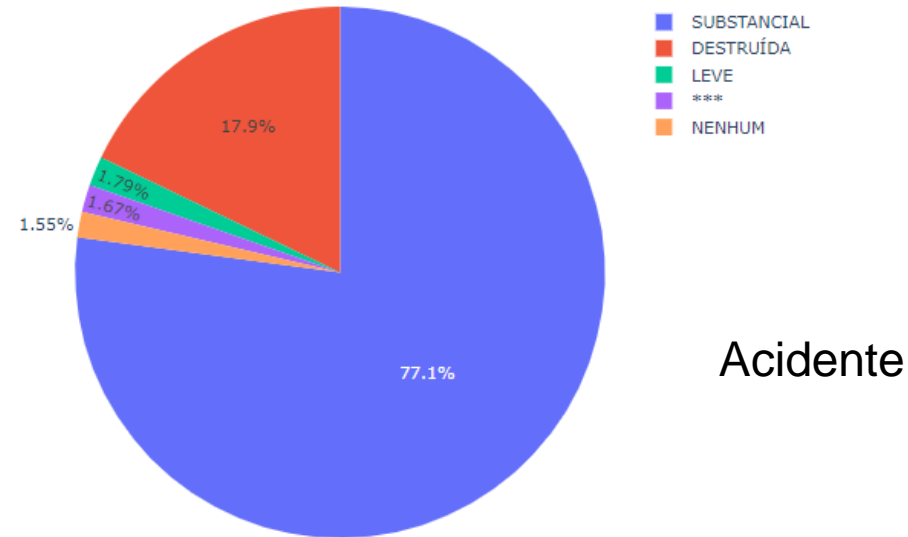
- Hipótese 1) Qual é a porcentagem de classificação das ocorrências relatadas?



A porcentagem de ocorrências de acidentes foi 32,4%(equivalente a 1667 ocorrências). O acidente representa justamente uma ocorrência mais significativa em termos de perdas materiais e humanas. Por esta razão, as análises apresentadas no restante deste trabalho consideram apenas os acidentes

ANÁLISE EXPLORATÓRIA

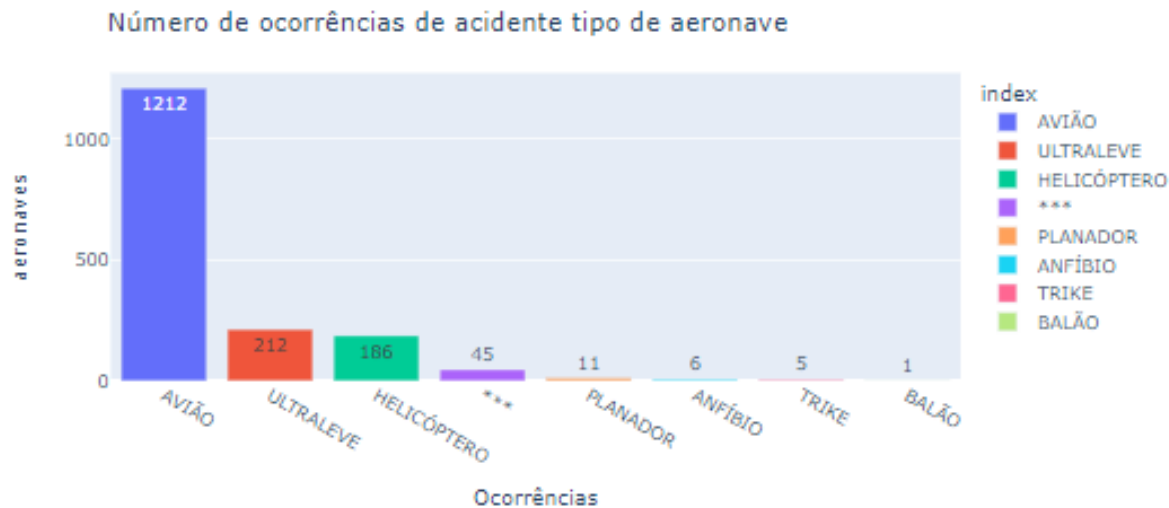
- Hipótese 2) Quais são os tipos de danos mais relatados nas ocorrências de acidente?



17,9% das 1667 ocorrências de acidente tiveram aeronaves destruídas e 77,1 com danos substancial. Cerca de 5% são ocorrência tiveram danos classificadas como leve, nenhum ou não identificada. Essa figura evidência a perda material das ocorrências com acidente.

ANÁLISE EXPLORATÓRIA

- Hipótese 3) Qual são os tipos de aeronaves e o tipo de motor mais relatadas nas ocorrências de acidente?

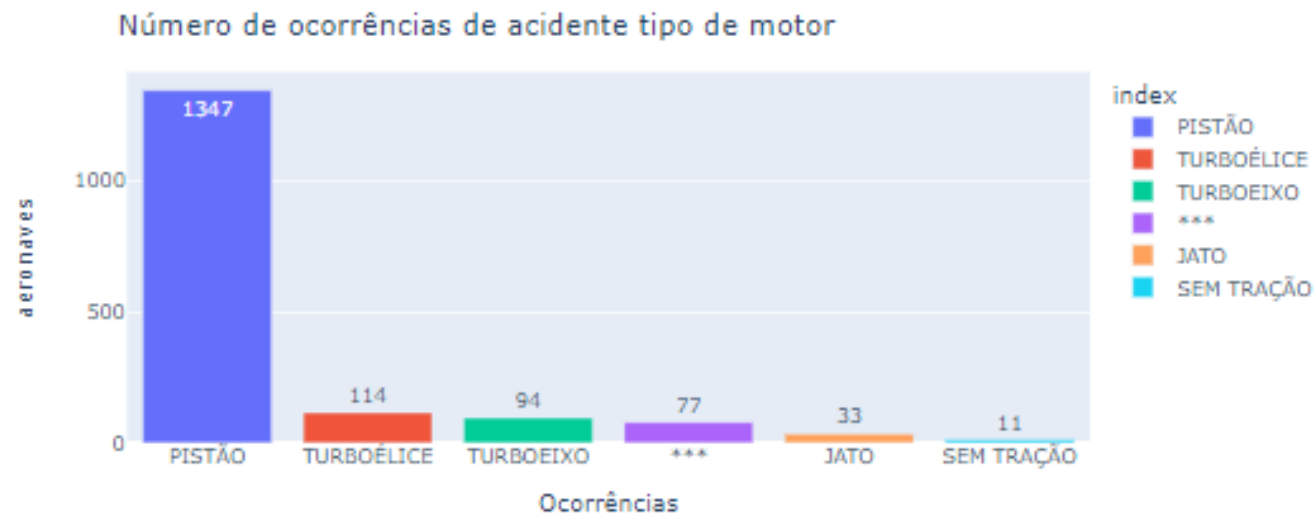


Acidente

O avião é um tipo de aeronave mais frequente em ocorrências aéreas no Brasil, 1212 das 1667 ocorrências de acidentes que foram relatadas em aviões. Seguido por Ultraleves 212 e Helicóptero 186 ocorrências.

ANÁLISE EXPLORATÓRIA

- Hipótese 3) Qual são os tipos de aeronaves e o tipo de motor mais relatadas nas ocorrências de acidente?

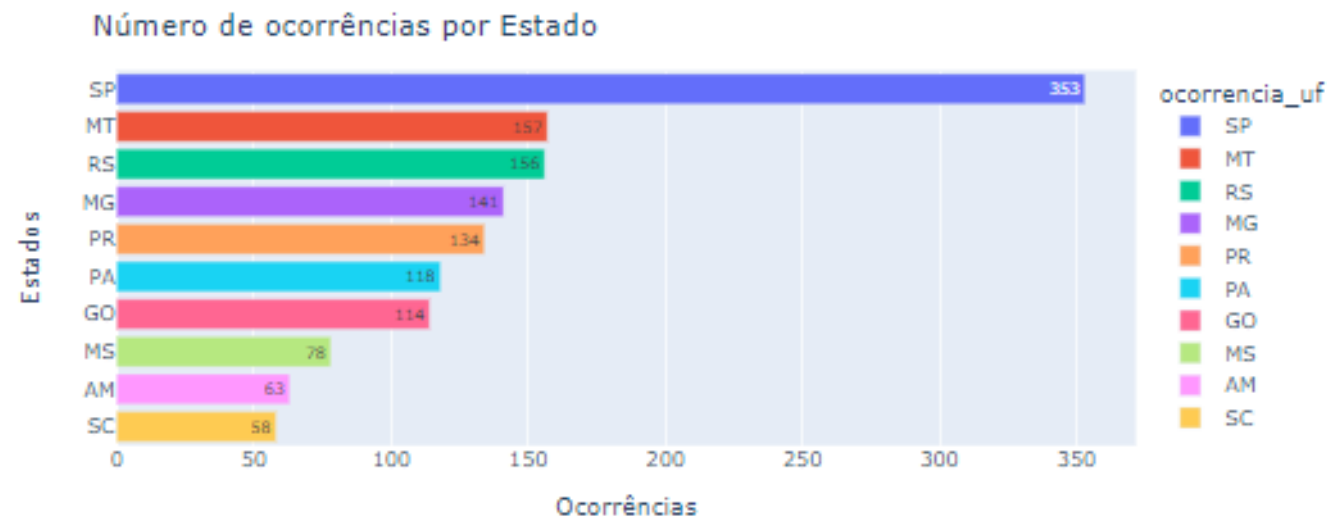


Acidente

Os motores aeronáuticos são classificados em motores a pistão, a jato, turboélice e os turboeixo. É possível observar, no gráfico que os motores a pistão são encontrados em maior frequência nas ocorrências de acidente reportadas. Este tipo de motor é o mais utilizado em aviões de pequeno porte ([Link](#)).

ANÁLISE EXPLORATÓRIA

- Hipótese 4) Quais são os estados mais relatados nas ocorrências de acidente?

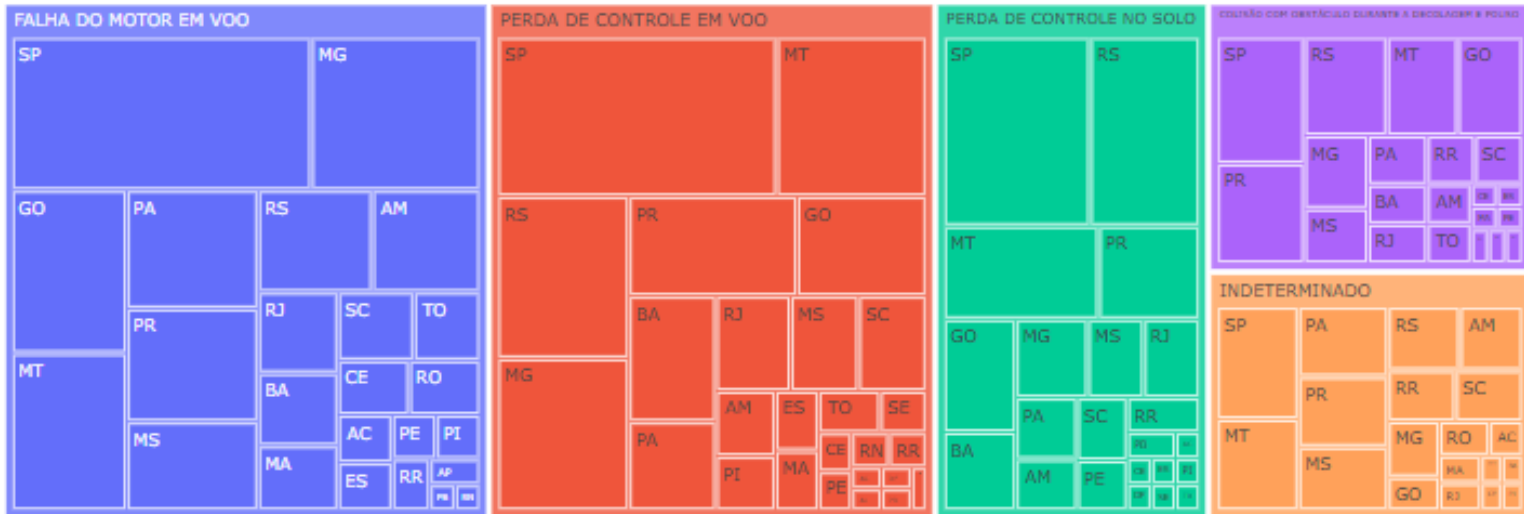


Acidente

A quantidade de registros de acidentes em São Paulo foi expressivamente maior que a quantidade de registros demais estados. Esse número é devido à concentração de aeródromos no Estado de São Paulo ([fonte](#)).

ANÁLISE EXPLORATÓRIA

- Hipótese 5) Quais são os tipo de ocorrências de acidente mais relatados?



Acidente

No treemap é possível observar o tipo de ocorrência por estado. As causas dos acidentes mais relatadas foram falha do motor em voo (SP e MG), perda de controle em voo (SP e MT), perda de controle no solo (SP e MG). As causas para tais ocorrências não necessariamente são em decorrência de uma falha humana, pois a perda de controle em solo ou em voo pode ser devido a uma falha estrutural, mecânica, ou elétrica na aeronave.

ANÁLISE EXPLORATÓRIA

- Hipótese 6) Qual é o número de ocorrências de acidente por ano?



Em relação a série temporal dos acidentes, ocorreu redução dos casos registrados entre os anos de 2012 (206 acidentes) à 2021(142 acidentes). Em 2018 ocorreu um ligeiro aumento.

PREDIÇÃO DA CLASSIFICAÇÃO DAS OCORRÊNCIAS

MODELAGEM DA CLASSIFICAÇÃO DAS OCORRÊNCIAS

Model	Accuracy	AUC	Recall	Prec.	F1	Kappa	MCC	TT (Sec)
Random Forest Classifier	0.9016	0.9810	0.8224	0.8971	0.8933	0.8366	0.8403	2.126
Extra Trees Classifier	0.9040	0.9806	0.8342	0.8997	0.8980	0.8413	0.8438	2.679
Logistic Regression	0.9075	0.9805	0.8457	0.9038	0.9033	0.8477	0.8494	6.158
Gradient Boosting Classifier	0.9009	0.9794	0.8261	0.8967	0.8940	0.8360	0.8391	19.963

Um bom modelo de classificação é um modelo que pode separar bem as duas classes; portanto, melhor será a AUC (*Area Under the Curve*).

MODELAGEM DA CLASSIFICAÇÃO DAS OCORRÊNCIAS

Model	Accuracy	AUC	Recall	Prec.	F1	Kappa	MCC	TT (Sec)
Random Forest Classifier	0.9016	0.9810	0.8224	0.8971	0.8933	0.8386	0.8403	2.126

Fold	Accuracy	AUC	Recall	Prec.
0	0.8741	0.9698	0.8256	0.8781
1	0.9016	0.9745	0.8457	0.8978
2	0.8998	0.9733	0.8585	0.9012
3	0.8929	0.9743	0.8379	0.8913
4	0.8981	0.9798	0.8390	0.8938
Mean	0.8933	0.9743	0.8413	0.8924
Std	0.0100	0.0032	0.0108	0.0079

RandomForestClassifier Confusion Matrix

True Class	ACIDENTE	362	1	21
	INCIDENTE	1	420	21
	INCIDENTE GRAVE	28	21	91
		ACIDENTE	INCIDENTE	INCIDENTE GRAVE
		Predicted Class		

CONSIDERAÇÕES FINAIS

- A análise exploratória dos dados permitiu traçar o perfil de ocorrências de acidentes, com base nos tipos de danos e aeronaves, localidades e horários de maior ocorrência, bem como o tipos de ocorrências. Assim é possível identificar tratativas afim de prevenir as ocorrências de acidente.
- Com base nas informações coletadas é possível utilizar o modelo de predição para classificar automaticamente as ocorrências em acidentes, incidentes graves e incidentes.