

# Ocorrências Aeronáuticas na Aviação Civil Brasileira

Héríkles Cordeiro

Trabalho teve como fim ser avaliação do teste técnico para a vaga de Cientista de Dados da A3Data base dados "Ocorrências Aeronáuticas na Aviação Civil Brasileira" dos dados abertos do governo (<https://dados.gov.br/dataset/ocorrencias-aeronauticas-da-aviacao-civil-brasileira>)

O objetivo desse trabalho é realizar uma análise exploratória com foco nas ocorrências com fatalidades, levantamento de Hipóteses/insight e , por fim, responder às perguntas levantadas no texto. Com o objetivo de identificar quais características estão relacionadas a esse evento.

Encarado como se o presente participante já estivesse integrado a equipe da A3Data e essa base de dados fosse apresentada por um novo cliente. Esta seria o equivalente a primeira entrega para o cliente.

## Metodologia do Processo Utilizado

Posto esta condição, inicialmente foi realizado um processo de investigação nos conjuntos de dados indicados nos testes, o intuito foi entender os dados disponíveis e definir um escopo inicial. Toda atividade feita nesse sentido foi descrita em outro documento `Analise_e_limpeza_dados.ipynb`, optou-se por essa metodologia para deixar esse presente documento apenas as análises de dados e storytelling direcionados para um possível cliente leigo. O outro documento é direcionado para discussões dentro da equipe de Dados.

### OCORRÊNCIA:

- \* codigo\_ocorrendia = Representará todos os id's;
- \* ocorrencia\_classificacao;
- \* ocorrencia\_uf;
- \* ocorrencia\_dia = será transformado em 'ocorrencia\_mes' e 'ocorrencia\_ano';
- \* ocorrencia\_hora = será transformado em horas completas;
- \* total\_aeronaves\_envolvidas;

OCORRÊNCIA\_TIPO

- \* codigo\_ocorrencia1 = será convertida para 'codigo\_ocorrencia';
- \* ocorrencia\_tipo;

## AERONAVE

- \* codigo\_ocorrencia2 = será covertida para 'codigo\_ocorrencia';
- \* aeronave\_registro\_segmento;
- \* aeronave\_tipo\_veiculo;
- \* aeronave\_fabricante;
- \* aeronave\_modelo;
- \* aeronave\_motor\_tipo;
- \* aeronave\_motor\_quantidade;
- \* aeronave\_assentos;
- \* aeronave\_ano\_fabricacao;
- \* aeronave\_nivel\_dano;
- \* aeronave\_fase\_operacao;
- \* aeronave\_fatalidades\_total;



### FATOR\_CONTRIBUINTE

- \* codigo\_ocorrencia3 = será covertida para 'codigo\_ocorrencia';
- \* fator\_area;

## Definição dos algoritmos usados para entender a base de dados

Como descrito anteriormente, o objetivo final desse Trabalho foi definido como encontrar possíveis relações entre as variáveis acima citadas e ocorrências com fatalidades. Para atingir essa meta, foram escolhidas duas técnicas de Aprendizado de Máquina com a capacidade de descrever as relações presentes na base de dados. As Técnicas são Regressão Logística e Árvores de Decisão, ambas as técnicas têm como características uma alta aplicabilidade.

# Hipóteses

1. Aeronaves mais antigas tem um maior acidentes com fatalidade?
2. Ocorrências com fatalidade têm predominância no período noturno?
3. Quais meses tem maior número de ocorrência?
4. Quais meses tem maior número de ocorrência com fatalidade?
5. Algum modelo está mais associado a ocorrência com fatalidade?
6. Algum Fabricante está mais associado a ocorrência com fatalidade?
7. O maior número de ocorrência com fatalidade está entre quais tipos aeronaves?
8. É possível apontar que a maioria das ocorrências com fatalidade está relacionado a fatores humanos?
9. Qual o resultado apresentado pelos algoritmos de ML?

# Hipóteses

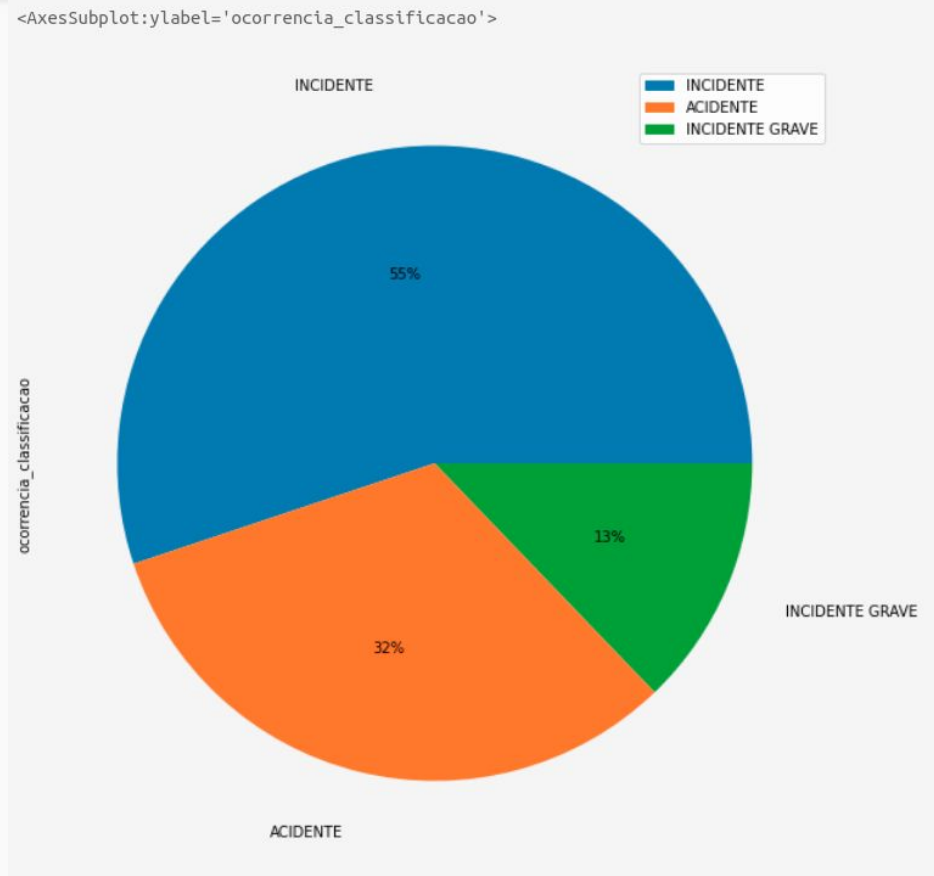
1. Aeronaves mais antigas tem um maior acidentes com fatalidade?
2. Ocorrências com fatalidade têm predominância no período noturno?
3. Quais meses tem maior número de ocorrência?
4. Quais meses tem maior número de ocorrência com fatalidade?
5. Algum modelo está mais associado a ocorrência com fatalidade?
6. Algum Fabricante está mais associado a ocorrência com fatalidade?
7. O maior número de ocorrência com fatalidade está entre quais tipos aeronaves?
8. É possível apontar que a maioria das ocorrências com fatalidade está relacionado a fatores humanos?
9. Qual o resultado apresentado pelos algoritmos de ML?

# Análise Exploratória

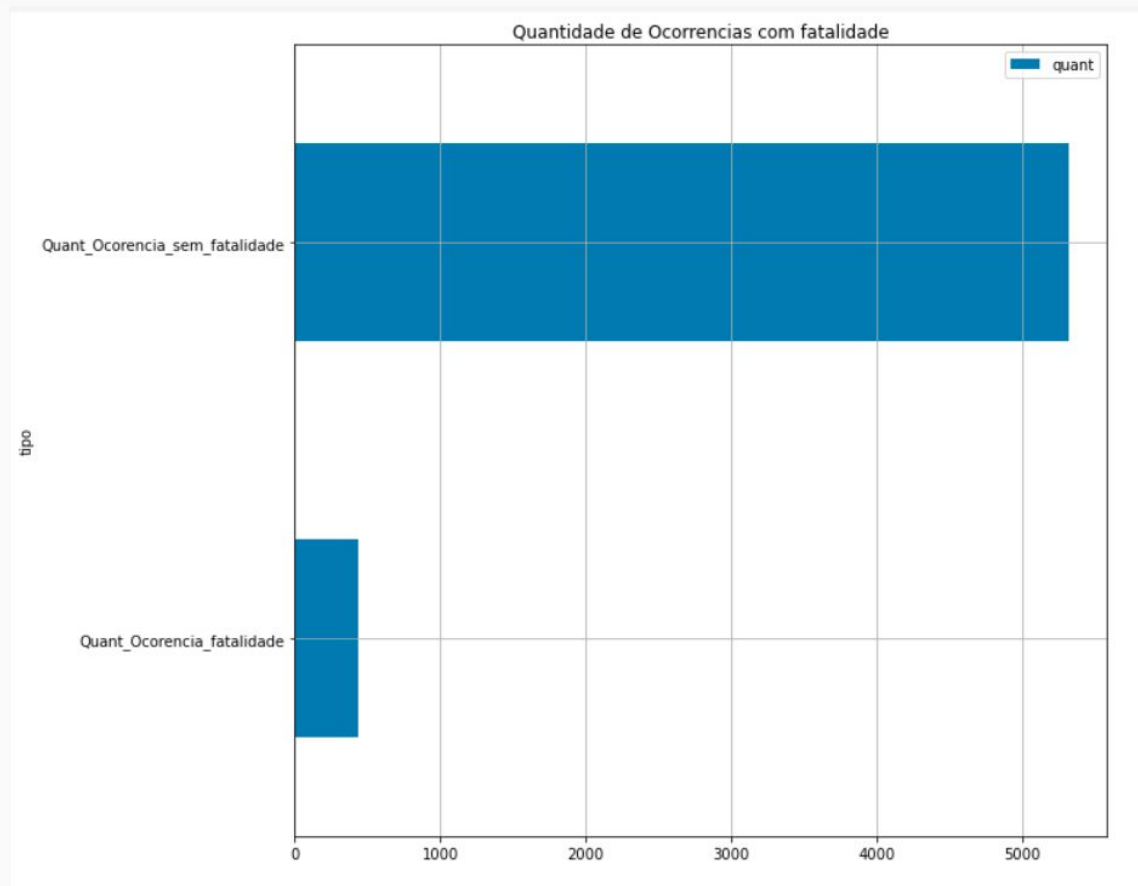
aeronave\_tipo\_veiculo

	quant	porc
AVIÃO	4683	80.45%
HELICÓPTERO	654	11.24%
ULTRALEVE	333	5.72%
***	110	1.89%
PLANADOR	19	0.33%
ANFÍBIO	13	0.22%
TRIKE	5	0.09%
DIRIGÍVEL	2	0.03%
HIDROAVIÃO	1	0.02%
BALÃO	1	0.02%

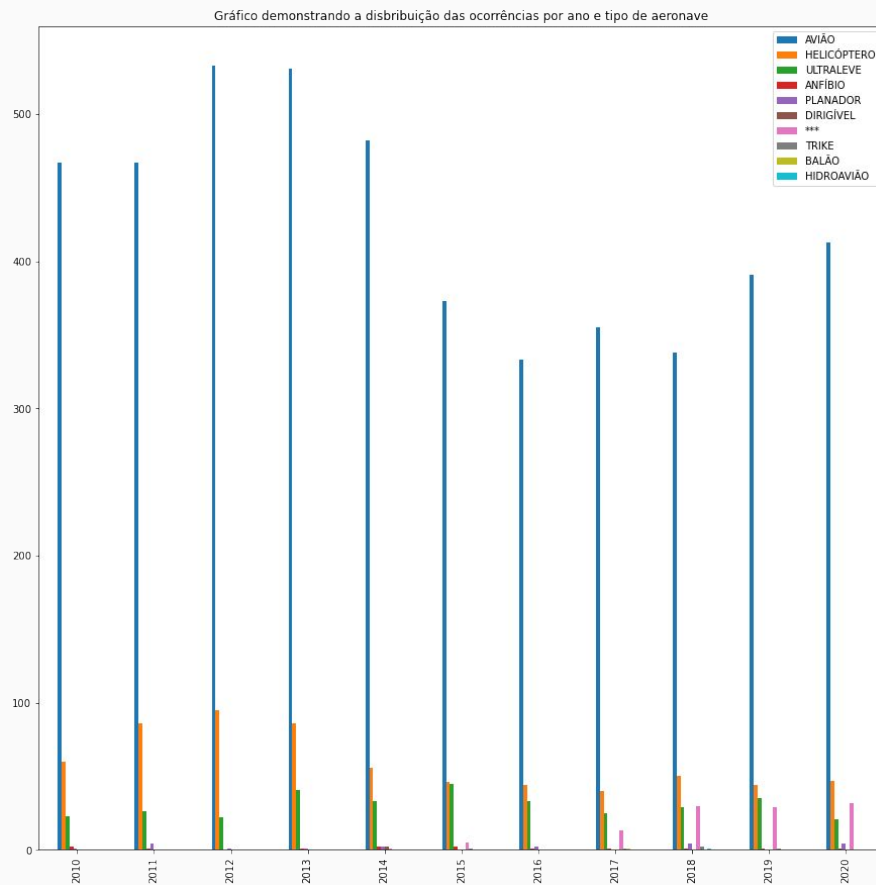
# Análise Exploratória



# Análise Exploratória

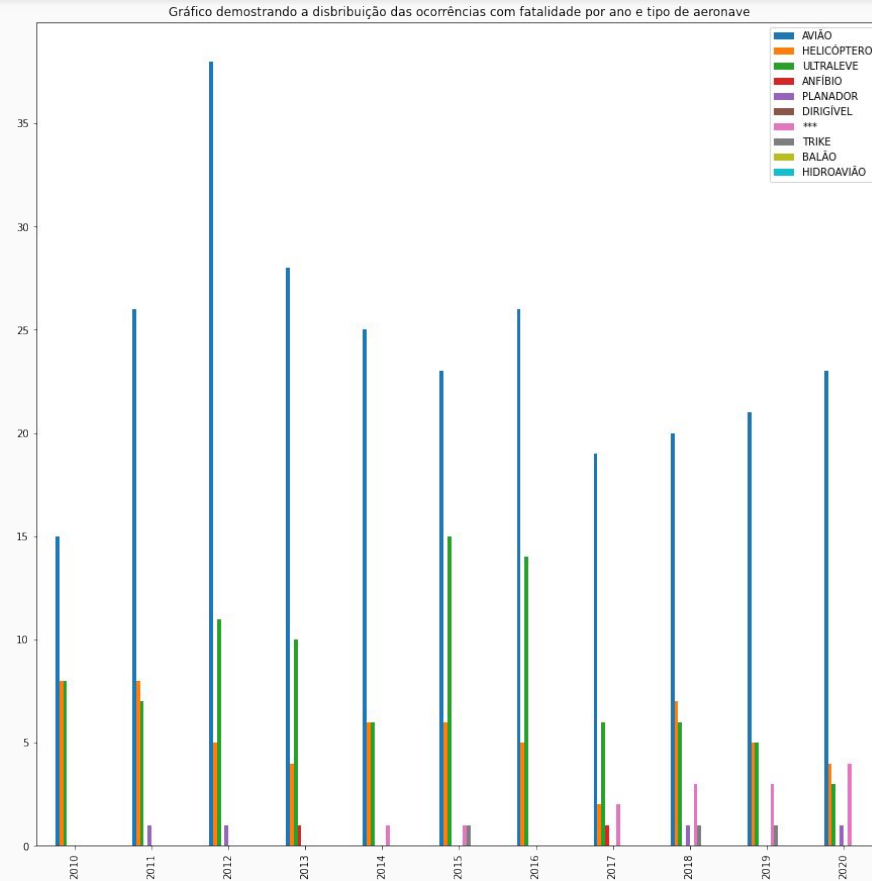


# Tipos de Aeronaves

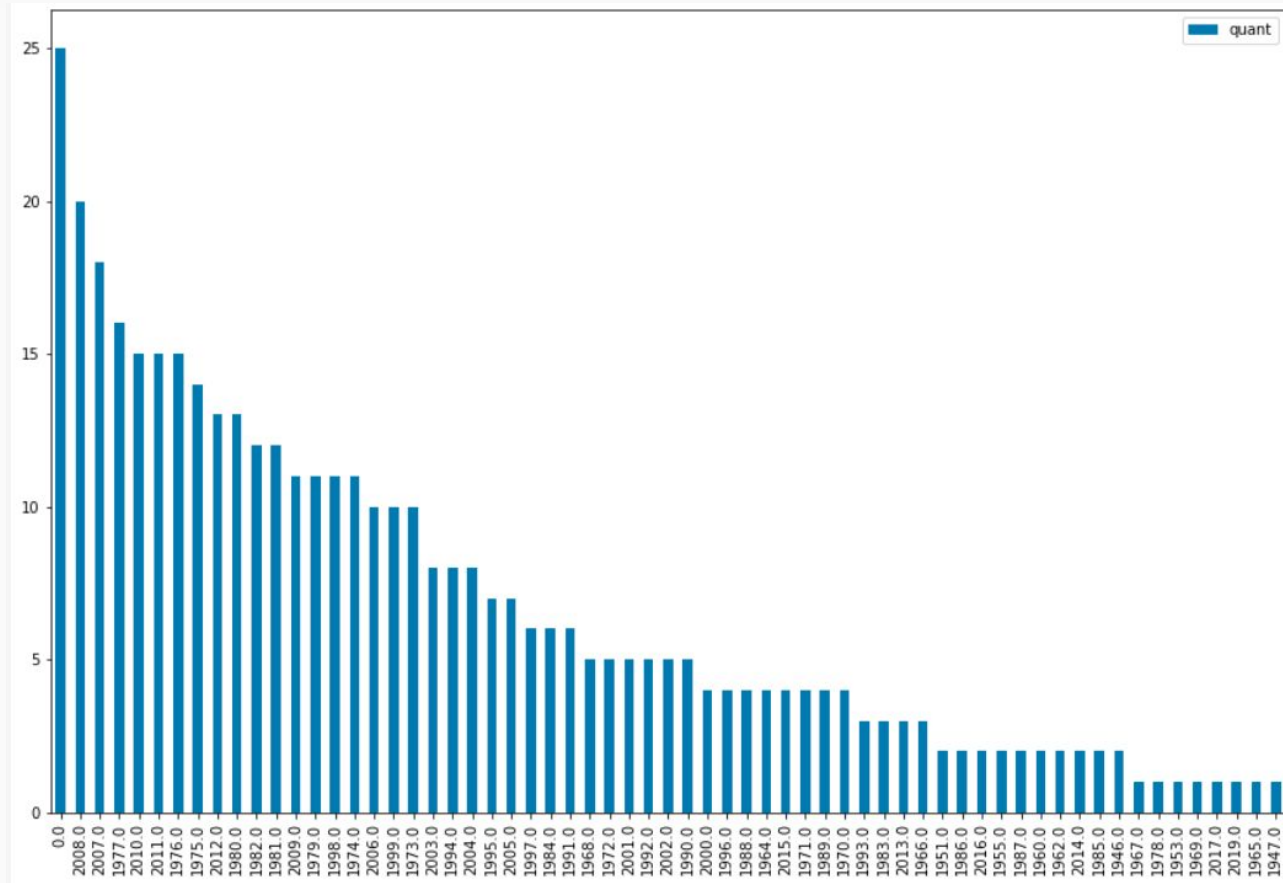




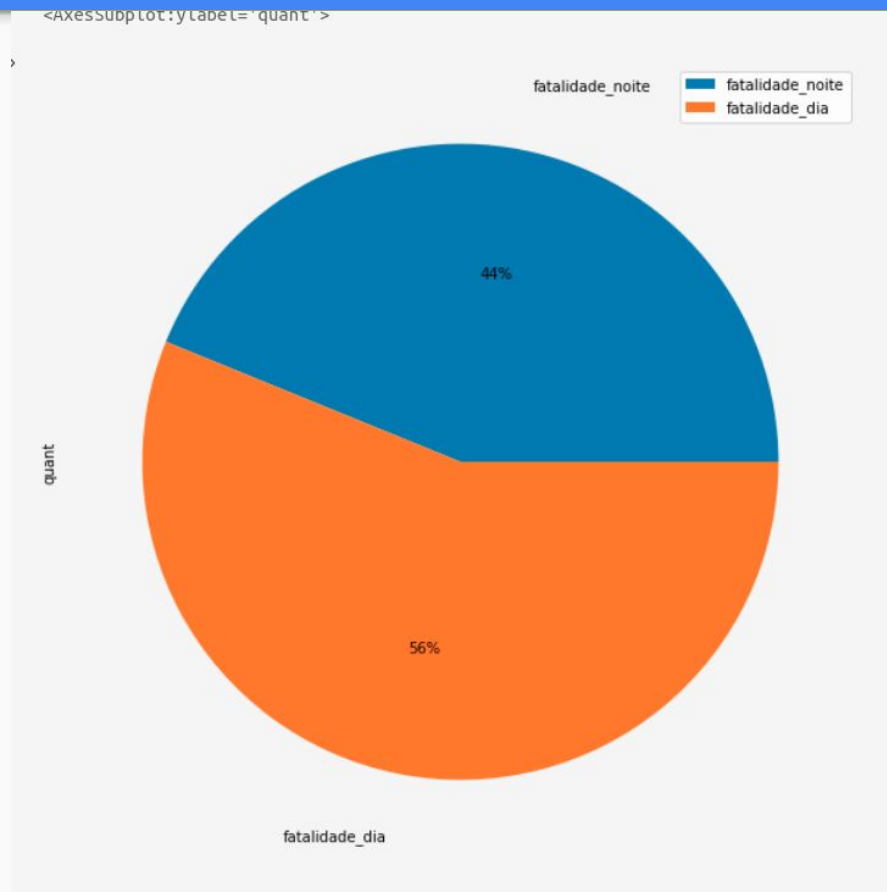
# Tipos de aeronaves com fatalidade



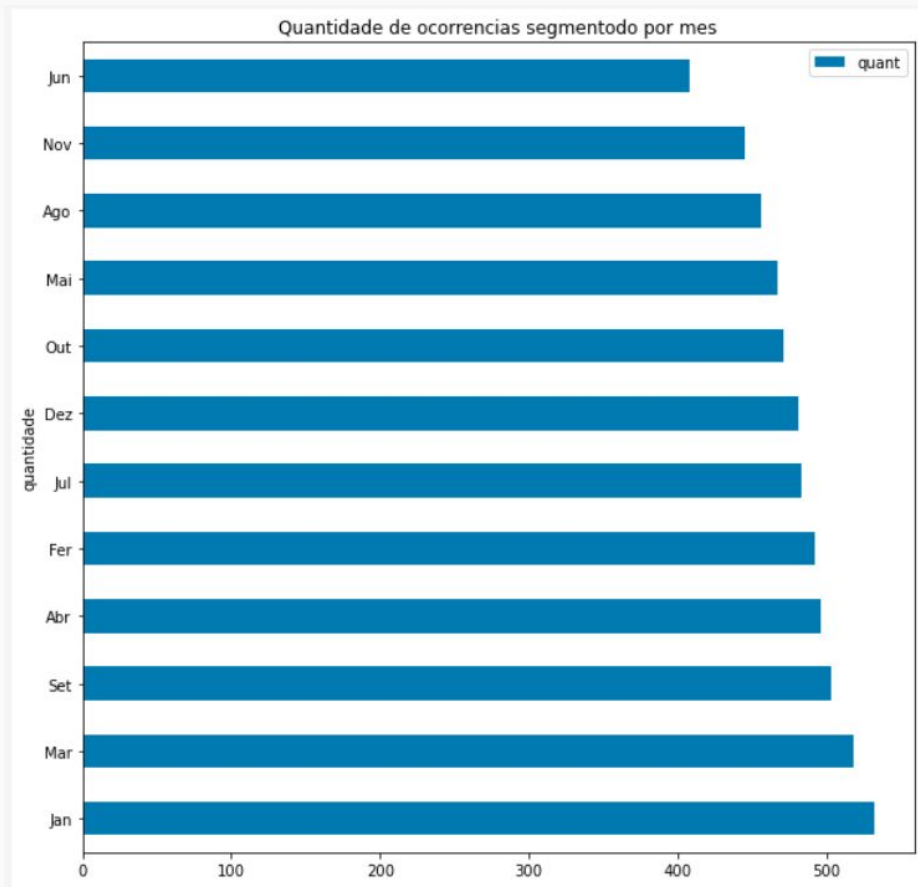
# Aeronaves mais antigas têm um maior número acidentes fatalidade?



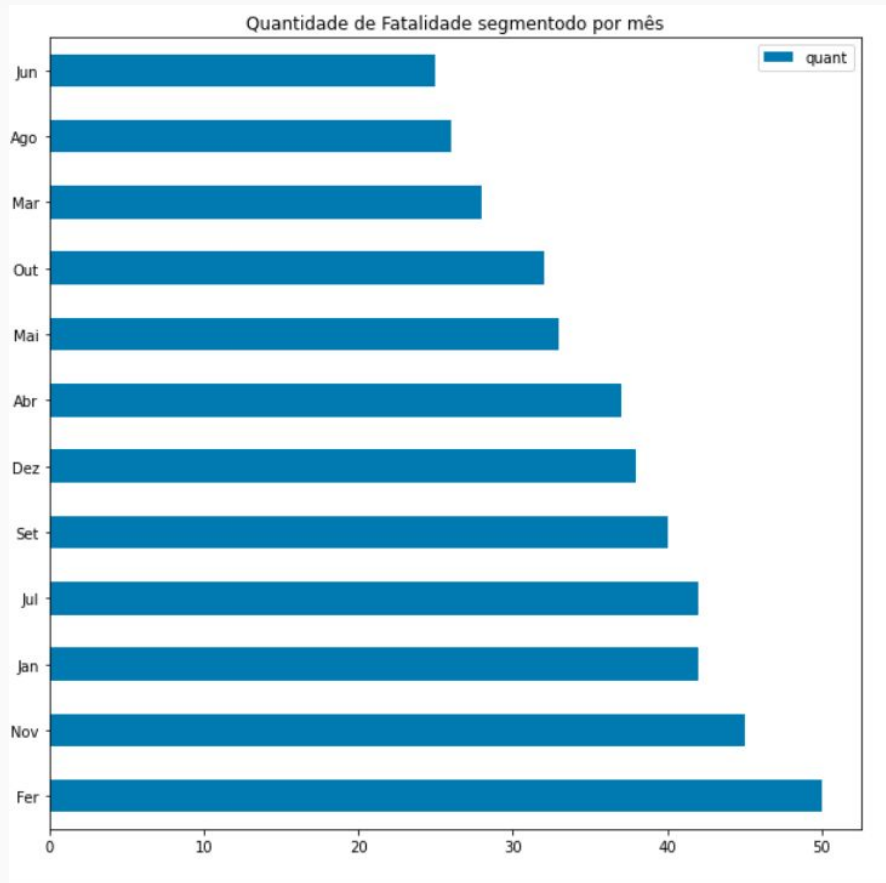
# Horário do dia que acontecem mais acidentes



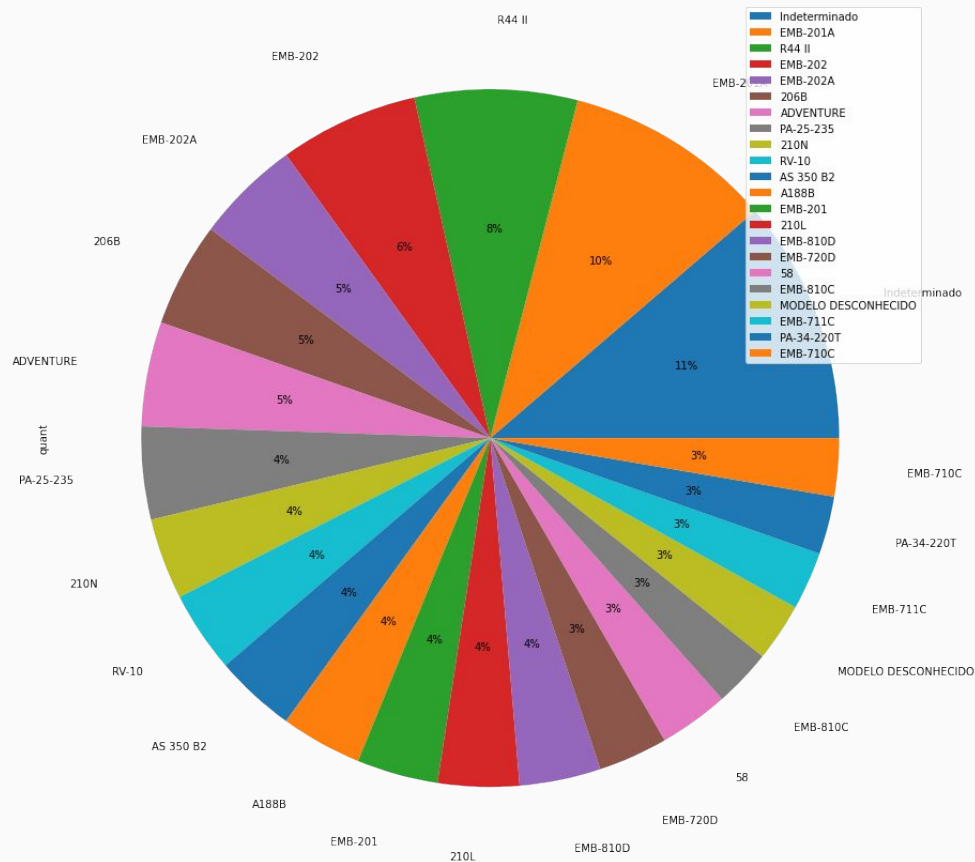
# Quantidade de ocorrências segmentada por mês



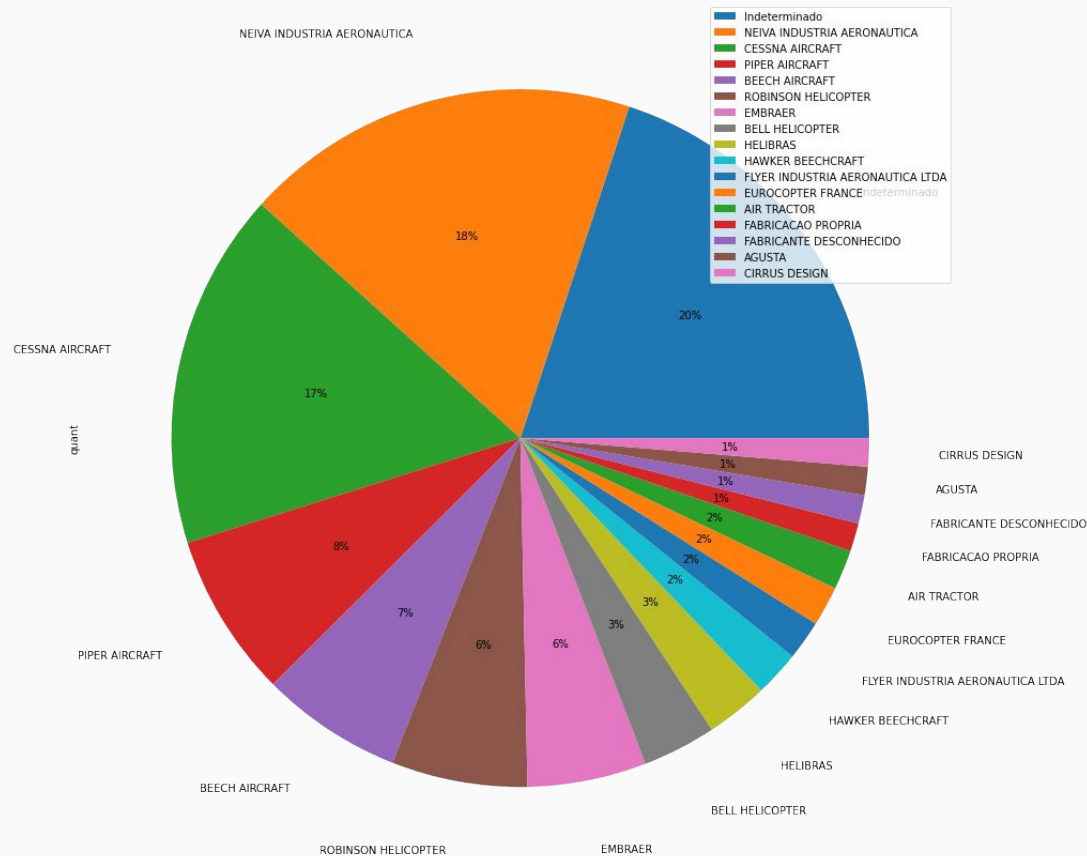
# Quantidade de ocorrências com fatalidade segmentada por mês



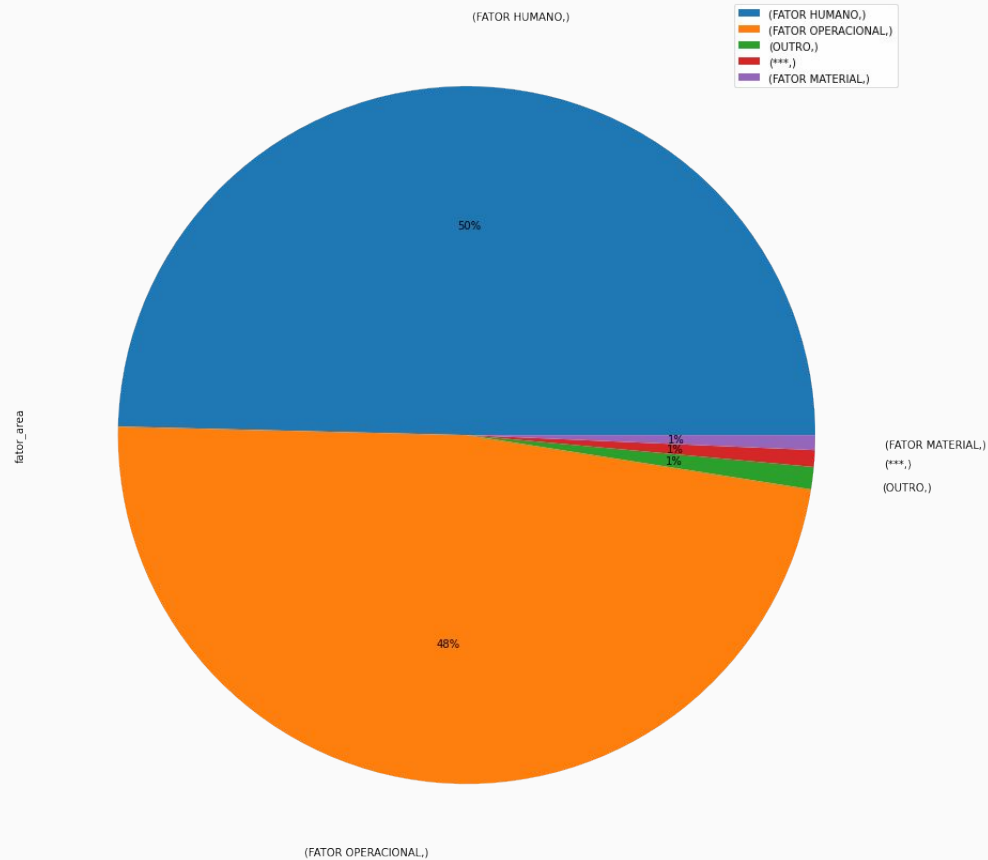
# Quantidade de ocorrência com fatalidade por modelo de aeronave



# Quantidade de ocorrência com fatalidade por Fabricante de aeronave



# Fatores causadores de Ocorrências





## Conclusão

Pelo fato do email contendo o desafio ter ido parar na caixa de spam, esse estudo foi executado em 1 dia. Por esse motivo a implementação dos algoritmos de aprendizado de máquina teve que ser interrompido.

Dadas as possibilidades de abordagem diferentes, presentes nessas base de dados, o caminho escolhido foi a avaliação e levantamento de hipóteses perante as ocorrências com fatalidade. As perguntas iniciais descritas na seção 3 foram respondidas:

# Conclusão

1. Aeronaves mais antigas têm um maior número de acidentes com fatalidade?

\* Não necessariamente, dentre as 5 datas com identificação, 4 tem menos de 20 anos.

2. Ocorrências com fatalidade têm predominância no período noturno?

\* Pelo contrário, a análise mostrou que cerca de mais da metade das ocorrências com fatalidade aconteceram de dia.

3. Quais meses tem maior número de Ocorrência?

\* Janeiro, Março e Setembro.

4. Quais meses tem maior número de Ocorrência com fatalidade?

\* Fevereiro, novembro e janeiro

5. Alguma modelo está mais associado a Ocorrência com fatalidade?

\* Os modelos com maior porcentagem são EMB\_201A, R44 II, e EMB-202. Porém, não traz nenhuma grande distância entre os demais, ainda que não exista essa associação entre modelos e Ocorrência com fatalidade.

# Conclusão

6. Algum Fabricante está mais associado a Ocorrência com fatalidade?

\* As fabricantes NEIVA INDUSTRIA AERONAUTICA e CESSNA AIRCRAFT apresentam elevado número de ocorrência comparada aos demais, indicando uma possível característica negativa perante as demais.

7. O maior número de ocorrência com fatalidade está entre quais tipos aeronaves?

\* Nas aeronaves do modelo Avião, tendo em número elevado de Ocorrência com e sem fatalidades. Porém o que deve ser ressaltado é que ~80% das Ocorrência registradas são desse tipo de modelo.

8. É possível apontar que a maioria das ocorrências com fatalidade estão relacionadas a fatores humanos?

\* Sim, conforme apresentado acima.

9. Qual o resultado apresentado pelos algoritmos de ML?

\* Não foi possível obter essa resposta, dados os imprevistos citados.

Como próximos passos, é indicado a realização dos testes dos modelos de aprendizado de máquina, e com os resultados decidir qual melhor caminho a se realizar. Outro ponto a ser reanalisado são algumas variáveis que foram, deixadas de fora da análise devido ao prazo inicial. E quando possível também executar testes estatísticos mais aprofundados.