**Observações**

* Tem disponível três anos do abono de permanência e aposentados
* O que é abono [permanência](https://ingracio.adv.br/abono-permanencia-servidor-publico/) ?
* Arquivo ABONO: quem está recebendo abono até aquela data de extração.
* Versão compilada: acredito que seja todos os aposentados até aquele momento

**Passo-a-passo**

Pré-processamento dos dados

* Limpeza de dados. Procurar por: valores faltantes, dados ruidosos, Outliers, Inconsistências. Valores faltantes: ignorar a tupla, algum valor global, estatística , probabilidade, criar uma nova coluna dizendo que teve dado faltante. Para os dados ruidosos aplicar Binning, regressão, Análise de Outliers
* Integração de dados. Reduzir a redundância e inconsistência dos dado com: identificação de entidade, análise de redundância e correlação, duplicação de tupla, detecção de conflito de valores de dados
* Redução dos dados. Tentar aplicar técnicas de redução de dimensionalidade.
* Transformação de dados. Verificar qual utilizar: smoothing, construção de atributos, agregação, normalização, discretização, geração de hierarquia.

Estatística

* Realizar a estatística descritiva em todas as colunas, respondendo sempre a essas questões: qual o tipo de dado? Qual a melhor forma de visualizar os dados? Qual a média, moda, mediana, variância, (gerar gráficos como o violinplot)?
* Fazer hipóteses e inferências. Criar possíveis hipóteses sobre o problema, fazer o teste. Verificar se as variáveis seguem uma distribuição normal. Matriz de correlação

Aprendizado de máquina

* Caso tenha necessidade, faça alguma transformação nos dados.
* Se ainda não tiver, gere várias tabelas tabelas com diferentes inputs
* Defina o conjunto de treino, teste.
* Defina uma validação cruzada
* Defina um algoritmo.
* Faça o treino.
* Capture os parâmetros.
* Repita até que todas as tabelas sejam testadas.
* Repita os processos acima com outros algoritmos.

Resultados

* Analisar todos os resultados da etapa anterior e tirar conclusões, se possível mostrar gráficos, e etc

IDEIAS

Tabela aposentados

* Nome e CPF  
  Vão ser utilizados para formar a chave estrangeira que fará a ligação com a tabela de abono
* Matrícula  
  Me parece ser uma coluna desnecessária visto que se trata de um identificador pessoal e essa coluna não existe na tabela abono
* Orgão  
  Verificar quantos órgãos tem, % de aposentados por órgão, Quais tem mais/menos aposentados? Qual órgão que gasta mais/menos com aposentadoria? As maiores aposentadorias estão relacionadas a algum órgão?Em qual órgão o aposentado ganha em média mais?
* Sigla Orgao  
  Coluna desnecessária ,tem redundância com a coluna orgão
* Orgao Vinculacao  
  Essa coluna parece ser um agrupamento dos órgãos, ou seja, está em uma dimensão maior. Ex: Instituições de ensino tem o valor:1500. como vários órgãos não tem órgão de vinculação(número) acho podemos descartá-lo
* Cargo emprego  
  Verificar quantos cargos tem, % de aposentados por cargo, quais tem mais/menos aposentados? Qual cargo recebe mais/menos com aposentadoria? As maiores aposentadorias estão relacionadas a algum cargo?
* Classe  
  “Patamares em que o cargo efetivo ocupado pelo servidor aposentado encontrava-se dividido, podendo ser subdivididos em padrões, quando existentes”. São colunas com valores numéricos e CHAR. Como não sei a origem desses dados e sua relevância, podemos transformar tudo para categórico e utilizar essa coluna
* Padrão

“Subdivisão que determina o posicionamento do servidor aposentado na estrutura remuneratória do

cargo.” Como a descrição cita uma possível vinculação com a remuneração, podemos tudo para categórico e analisar essa coluna

* Referência

“Patamares em que o cargo efetivo que era ocupado pelo servidor aposentado encontrava-se dividido

quando se deu o ato de aposentadoria.”

Todos os valores são 0, acredito que não seja necessário utilizá-lo

* Nível

“Nível em que determina o posicionamento do servidor aposentado na estrutura remuneratória do cargo.”

Não faço ideia do impacto dessa variável

* Tipo de aposentadoria

Acredito que seja interessante. QUais são os tipos de aposentadorias? tem alguma correlação o tipo com o valor recebido? **pesquisar sobre aposentadoria por invalidez** podemos fazer algo por eles? tem como manter no trabalho?

* Fund Legal Aposentadoria

“Classificação sistêmica do tipo de aposentadoria conforme a legislação..” QUal a classificação que mais gera aposentadoria?

* Portaria aposentadoria

“Legislação que fundamentou a aposentadoria.”

Desnecessário

* Data da ocorrência

Buscar os aposentados mais novos, talvez os mais novos tenham interesse no abono. Qual a média de duração que a pessoa recebe a aposentadoria? ex: quantas pessoas estão recebendo a 10 anos a aposentadoria?

* Nome Ocorrência

Desnecessaŕio

* Data ingresso serviço público

Quanto tempo a pessoa passou no serviço público?

* Valor da aposentadoria

Média, moda, mediana, distribuição, outliers e etc...