

1) Considerando duas listas de strings em que cada lista possua valores repetidos calcule o ângulo vetorial através do cosseno, conforme equação abaixo:

O cosseno de dois vetores não zerados podem ser derivados utilizando a fórmula do produto dos pontos. A partir de dois vetores n-dimensionais, A e B, a similaridade do cosseno, $\cos(\theta)$, é representado utilizando o produto dos pontos e a magnitude dos vetores como:

$$\text{similarity} = \cos(\theta) = \frac{\mathbf{A} \cdot \mathbf{B}}{\|\mathbf{A}\| \|\mathbf{B}\|} = \frac{\sum_{i=1}^n A_i B_i}{\sqrt{\sum_{i=1}^n A_i^2} \sqrt{\sum_{i=1}^n B_i^2}},$$

Obs: Será necessário importar a biblioteca math -> import math

2) De maneira iterativa até que determinada condição seja alcançada, capture determinado conteúdo informado pelo usuário através do teclado e armazene em um arquivo.

3) Crie em memória uma matriz de valores aleatórios utilizando NumPy e DataFrame e depois armazene a matriz em um arquivo. Após isso carregue a matriz salva e calcule o seu traço de maneira iterativa, ou seja, itere sobre linhas e colunas para calcular o traço.

4) Crie um conjunto de dados com 3 colunas. Após isso crie mais uma coluna e incorpore ao conjunto de dados inicial. De maneira iterativa, indique determinada linha e coluna para ser recuperada até que determinada condição seja informada para finalizar a iteração.

5) Crie um arquivo em formato JSON e carregue o mesmo para um objeto. Após isso, de maneira iterativa informe determinado caminho para recuperar valores que constam no objeto JSON até que uma condição específica seja atingida.

- 6) Crie um conjunto de dados (DataFrame) contendo duas colunas. Depois crie mais um conjunto de dados com duas colunas em que a primeira coluna tenha os mesmos valores da primeira coluna do primeiro conjunto. Após isso, vincule/integre a segunda coluna do segundo conjunto ao primeiro conjunto levando em consideração as primeiras colunas dos dois conjuntos. Por fim, realize algumas recuperações de dados considerando uma e mais de uma coluna.**
- 7) Crie um conjunto de dados de maneira randômica com algumas centenas de linhas e algumas colunas. Analise o conjunto de dados resultante e defina um número. A partir disso modifique o conjunto de dados de modo que em nenhuma das colunas o valor mínimo e máximo seja excedido ao número definido. Ao final apresente as informações estatísticas do conjunto de dados para todas as colunas (contagem, média, mínimo, máximo, etc).**
- 8) Carregue o arquivo movie.csv. Após isso para cada valor distinto na coluna genres (que devem ser separados cada vez que o caracter | ocorre) gere novas colunas e incorpore ao conjunto original atribuindo 1. Após isso recupere o conteúdo de um intervalo de linhas do conjunto de dados.**
- 9) Crie um conjunto de dados com dois níveis hierárquicos de linhas e colunas. A partir do conjunto de dados realize operações de soma, média, mínimo e máximo considerando os níveis de hierarquia das linhas e colunas.**
- 10) Crie dois conjuntos de dados e realize o merge dos conjuntos. A partir do conjunto resultante efetue algumas filtrações nos dados considerando uma e mais de uma coluna.**
- 11) A partir de um conjunto de dados realize a agregação de duas colunas em que os valores da primeira coluna serão as linhas e os valores da segunda coluna as colunas. Para os valores da célula utilize uma terceira coluna e uma operação de soma. O conceito de tabela de pivotamento deve ser utilizado. Por fim, os dados da tabela resultante deve ser plotados em um gráfico de barras.**

12) Considerando o conjunto de dados 2015_USPTOf.json (patentes americanas) realize as devidas transformações para que cada um dos valores que fazem parte do atributo "Subclass_labels", visto que o atributo é um *array*, se torne uma coluna que deve ser adicionada ao conjunto de dados original. Além disso, considerando determinada linha, as colunas que são identificadas pelos valores distintos do atributo "Subclass_labels" devem ter o valor 1 atribuído. Ou seja, somente a coluna que se refere a determinada classe de uma patente deve ser preenchida com 1. Ao final realize o somatório de cada uma das colunas adicionadas (referentes aos valores do atributo "Subclass_labels") e apresente um histograma através de um gráfico de barras horizontal partindo das classes que mais ocorrem para àquelas que menos ocorrem. Apresente somente as 10 primeiras classes.

13) "DESAFIO" : Utilizando alguma API disponibilizada por redes sociais, tais como, YouTube, Instagram, LinkedIn, Facebook, entre outras, obtenha alguns dados públicos, ou seja, somente dados que estejam disponíveis via API. A partir disso, realize as devidas organizações para que os dados coletados seja estruturados em um DataFrame do Pandas. Com os dados estruturados utilize alguma função de agregação (soma, média, contagem) projetando o resultado de maneira gráfica.

In []: