

Ciência de dados com Python: nesta tarefa, irão explorar conjuntos de dados, operações para o seu processamento e visualização.

Criem uma conta no Github e criem um repositório para cada bloco de respostas, onde devem colocar todos os ficheiros (Readme, dados, Jupyter Notebook, doc, pdf, etc). Cada grupo partilha comigo o seu URL, para ir acompanhando o trabalho e efetuar a avaliação. No fim, fazem um ZIP com tudo e “colocam” no Nonio.

A – 2 valores

Escolham um conjunto de dados nos imensos repositórios disponíveis (dados abertos, públicos), sobre o tema que entenderem. O único requisito é que esse conjunto de dados possa ser “puxado” (download) como um arquivo CSV (valor separado por vírgula).

Respondam as seguintes questões num (novo) Jupyter Notebook:

- 1) O URL para o conjunto de dados;
- 2) Fundamentem a escolha deste conjunto de dados;
- 3) Escrevam um explicação do conjunto de dados, um Dicionário dos dados, que deve ficar no Readme do repositório.
- 4) Escrevam o código que carrega o conjunto de dados usando a função **read_csv** e mostrem as 10 primeiras linhas do conjunto de dados.
- 5) Qual é o tipo de dados de cada atributo?

Atenção: “tudo” tem de ser bem documentado, ie com comentários adequados (texto, imagens, etc), ao longo do Notebook.

B – 5 valores

Conjuntos de dados e IPYNB (Jupyter Notebook) em https://github.com/jgorvalho/PA2_2019-2020_TPC2

Questões e respostas (devidamente fundamentadas com comentários) no Jupyter Notebook: **tpc2_tarefa2.ipynb**

C – 9 valores

Visualização de Dados ...

C.1 – 5 valores

Criar um Jupyter Notebook para responder aos 4 exercícios, a seguir descritos. Usar a conjunto de dados referentes às vendas de uma empresa: `company_sales_data.csv`

Exercício 1: (1 valor) Ler o lucro total (Total Profit) de todos os meses e mostre-os através de um gráfico de linhas.

Usar o dado fornecido do lucro total para cada mês. A gráfico de linha gerado deve incluir as seguintes propriedades:

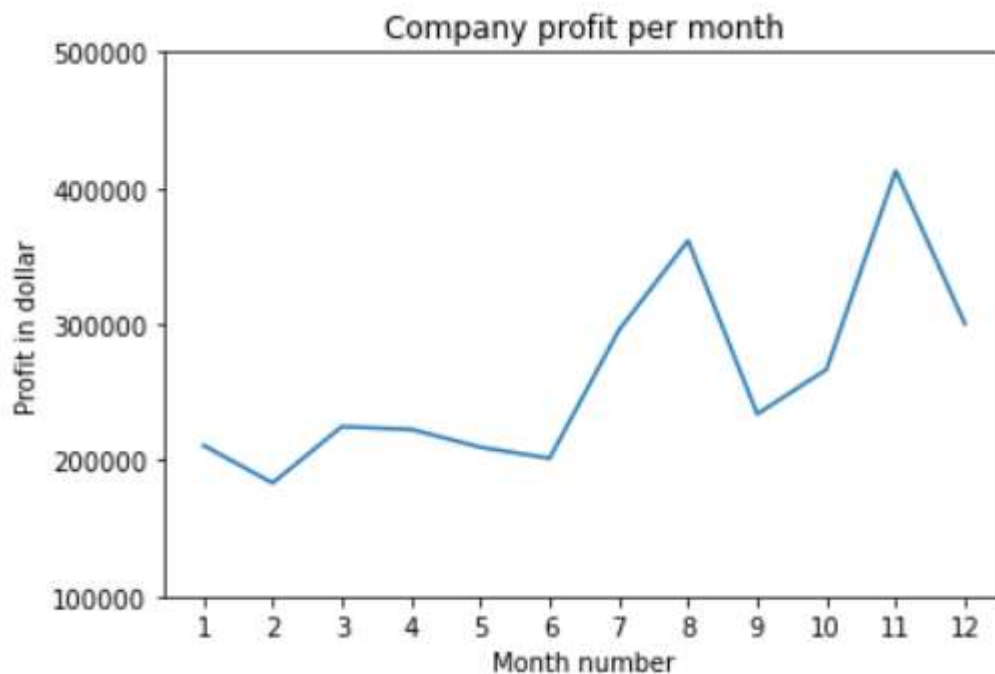
X label name = Month Number

Y label name = Total profit

Criar as seguintes listas:

```
profitList = df['total_profit'].tolist()
```

```
monthList = df['month_number'].tolist()
```



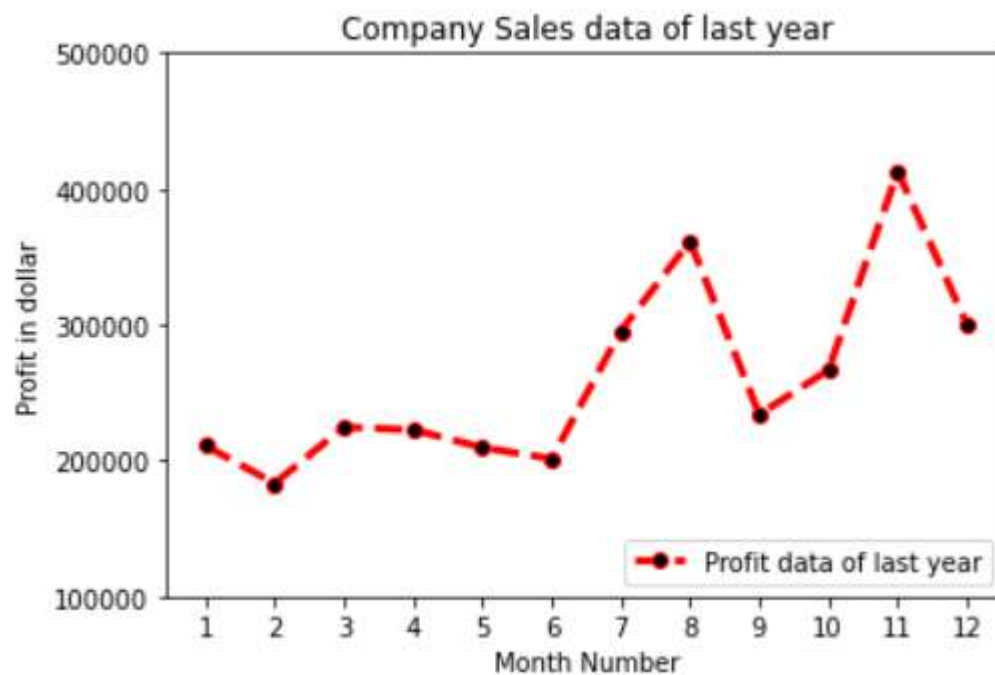
Exercício 2: (1 valor) Ler o lucro total (Total Profit) de todos os meses e apresente-os num gráfico de linhas, com as características indicadas.

A linha gerada deve incluir as seguintes propriedades de estilo:

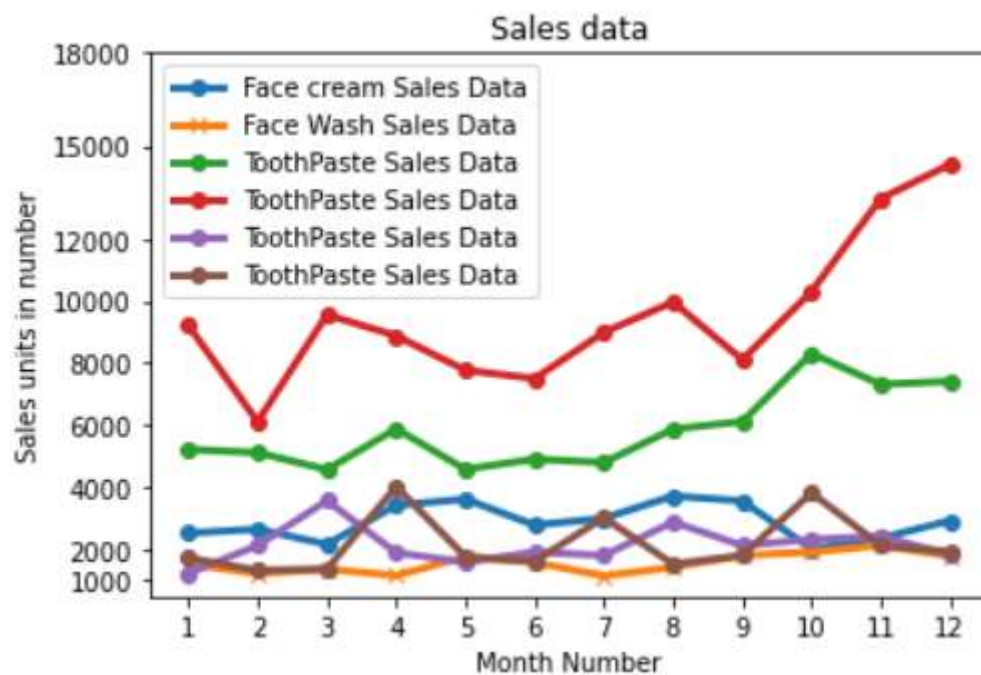
- Estilo de linha pontilhado e cor da linha devem ser vermelhos
- Mostrar legenda no local inferior direito
- Label do rótulo X = Número do mês
- Label do rótulo Y = Número de unidades vendidas
- Adicione um marcador de círculo
- Cor do marcador de linha conforme lida
- A largura da linha deve ser 3

Usar as listas:

```
profitList = df ['total_profit'].tolist()
monthList = df ['month_number'].tolist()
```

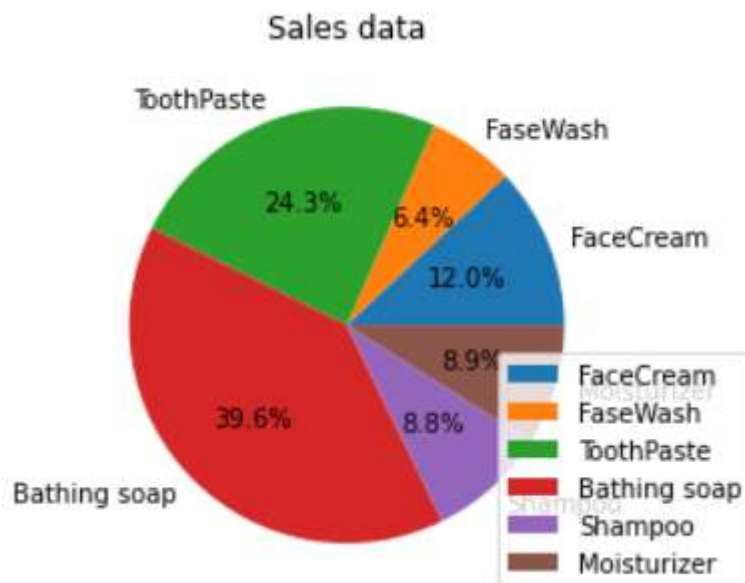


Exercício 3: (1 valor) Mostrar as vendas mensais de todos os produtos num gráfico de múltiplas linhas, ou seja, uma linha diferente para cada produto.



Exercício 4: (2 valores) Calcular para cada produto o total das vendas do ano e mostre-os usando Pie Chart (gráfico de queijo).

No Pie chart inscrever a % de vendas de cada produto.



C.2 – 4 valores

A palestra do TED

(https://www.ted.com/talks/hans_rosling_the_best_stats_you_ve_ever_seen#t-82288) de Hans Rosling sobre o desenvolvimento de país é um extraordinário de DataViz (Visualização de Dados). Mostra a evolução da esperança de vida e do PIB per capita, de vários países, através de uma animação. É uma oportunidade para conhecerem a Fundação Gapminder (www.gapminder.org).

Uma reprodução deste trabalho, usando Python, está neste programa:

gapminder_animation.p e neste Jupyter Notebook (Python, NumPy, Pandas, Matplotlib e Seaborn): gapminder_animation.ipynb

Atenção ao comentário final: é preciso instalar o Image Magick - <https://imagemagick.org/>

Numa versão diferente, está uma animação criada com a função FuncAnimation do Matplotlib, neste Jupyter Notebook: gapminder_animation-FuncAnimation_PLOT_SEM-legend.ipynb

Questão A: (1 valor) – Explique a metodologia para criar a animação em cada um dos casos (versões). Comente-as. Justifique.

Questão B: (1 valor) – Temos “bolhas” de cinco (5) cores diferentes. O que representa cada cor? Justifique, com descrição detalhada do processo.

Questão C: (2 valor) – Inserir uma legenda ou legendas com a cor das “bolhas” e o seu significado, posicionada(s) em área(s) que não se sobreponha(m) às “bolhas”. Fundamente as opções tomadas.