

Atividade - Python para CD e IA

- Nome/RA:
- Nome/RA:

Instruções

- Atividade pode ser feita em duplas;
- A entrega deve conter o arquivo no formato *.ipynb (Notebook Python) e em *.PDF do mesmo notebook!
- Prazo para entrega: **22/04/2023**

Problema: um estudo sobre o Câncer.

Contexto

A partir dos dados gerados por instituições no Estado de São Paulo que, sob coordenação da FOSP, mantém o Registro Hospitalar de Câncer (RHC) implantado, forma-se a base de dados estadual do RHC, aqui disponibilizada.

O principal objetivo da Unidade de Epidemiologia da FOSP ao disponibilizar este Banco de Dados é o de ampliar o acesso à base estadual e poder oferecer aos usuários, em especial técnicos da área da saúde, a possibilidade de execução de tabulações e análises específicas, de acordo com seu interesse e necessidade.

Vale ressaltar que a cada três meses este Banco de Dados é atualizado uma vez que a sistemática implantada prevê que trimestralmente os hospitais enviem à FOSP os dados referentes a casos novos diagnosticados e seguimento dos já cadastrados. Desta forma, os arquivos aqui disponibilizados, serão atualizados com a mesma periodicidade.

Alguns outros pontos merecem destaque:

- O Registro Hospitalar de Câncer do Estado de São Paulo cadastra casos novos diagnosticados a partir de 01/01/2000;
- Consideram-se casos analíticos para o RHC aqueles que chegaram à Instituição sem tratamento (podendo estar diagnosticados ou não) e são estes os casos contidos no arquivo aqui disponibilizado.
- Os dados coletados pelo RHC não devem e nem podem ser utilizados para cálculo de incidência, uma vez que retratam apenas o perfil de atendimento de uma determinada instituição (ou de um grupo destas).
- O arquivo aqui disponibilizado não identifica pacientes e nem as Instituições em que foram atendidos, sendo possível, porém executar tabulações segundo município de residência.
- Os tumores pediátricos diferenciam-se quanto à localização e morfologia em relação aos tumores em adultos e podem ser analisados segundo a Classificação Internacional do Câncer na Infância (CICI), que estabelece doze grandes agrupamentos de diagnóstico.

▼ Problema

Devemos fazer um estudo sobre o Câncer usando como referência o Banco de Dados mantido pela Fundação Oocentro de São Paulo (FOSP).

Nesse estudo, devemos estudar dois tipos de cânceres importantes: **colo do útero e câncer de mama**.

O Banco de Dados a ser utilizado pode ser encontrado no endereço:

<https://fosp.saude.sp.gov.br/fosp/diretoria-adjunta-de-informacao-e-epidemiologia/rhc-registro-hospitalar-de-cancer/banco-de-dados-do-rhc/>

Use o arquivo **Banco_RHC_referente_ao_ano_de_2022**.

EX01

O arquivo está no formato ***.dbf**.

Elabore uma rotina para a leitura e transformação desse formato ***.dbf bold text** para ***.csv**.

Após, crie um dataframe com esse arquivo ***.csv**.

1 # resposta

EX02

Pesquise os CIDs para colo do útero e câncer de mama. Crie um DF apenas com esses dois tipos de cancer. Dica: faça um filtro na coluna **TOPOGRUPC,3**.

1 # resposta

EX03

Faça um tratamento das células em branco e/ou que contenham valores inconsistentes.

1 # resposta

EX04

Utilizando o dicionário de dados disponível no Open LMS, resposta:

1. Quantos pacientes não tratados morreram de câncer em 2022?
2. Quantos pacientes não tratados morreram no total de 2022?
3. Quantos pacientes tratados/em tratamento morreram de câncer em 2022?
4. Quantos pacientes tratados/em tratamento morreram no total em 2022?

Faça a contabilização passo-a-passo.

Dica: trabalhe apenas com as colunas necessárias para responder as perguntas!

1 # resposta

EX05

Utilizando o dicionário de dados disponível no Open LMS, resposta:

1. Classifique, em ordem decrescente, os tipos de tratamento de câncer em 2022. Exiba esses resultados na forma numérica e gráfica.
2. Classifique, em ordem decrescente, os tipos de tratamento de câncer realizados no hospital em 2022. Exiba esses resultados na forma numérica e gráfica.

Faça a contabilização passo-a-passo.

Dica: Colunas TRATAMENTO e TRATHOSP!

1 # resposta

EX06

Utilizando o dicionário de dados disponível no Open LMS, resposta:

1. Existe alguma correlação entre a data do diagnóstico e o total de óbitos? Justifique sua resposta com uma ou mais análises.
2. Existe alguma correlação entre a data de início de tratamento e o total de óbitos? Justifique sua resposta com uma ou mais análises.

Faça a contabilização passo-a-passo.

Dica: Colunas DTDIAG e DTTRAT!

1 # resposta

EX07

Utilizando o dicionário de dados disponível no Open LMS, resposta:

1. Quais foram as 3 principais razões para a não realização do tratamento em 2022? Justifique.
2. Totalize, por cidade, o números de casos e óbitos pelos dois tipos de cânceres em 2022.

Faça a contabilização passo-a-passo.

EX08

Desafio: repita todos as análises anteriores utilizando o arquivo **Banco_RHC_completo_(até_12/2022)**.

Não há necessidade de apresentar o passo-a-passo! Aprensente os resultados apenas!

1 # resposta

▼ Dicas:

- Para análise gráficas, utilize as bibliotecas `matplotlib` e `seaborn`;
- Organize as respostas de forma objetiva;
- Faça um código comentado, de preferência, linha a linha;
- Esta análise é realizada com dados reais, portanto, as justificativas devem conter um rigor analítico;
- Tentem, para geração de códigos, ferramentas como Chat GPT. É uma excelente ferramenta para auxiliar em regras de sintaxe e para resolver problemas de erros no algoritmo;
- Utilizem as documentações das bibliotecas das funções. Somente assim, você conseguirá se ambientar com o ambiente de desenvolvimento Python.

1