|  |  |
| --- | --- |
| **Fonte de dados para a construção do documento:**  <https://dados.gov.br/dados/conjuntos-dados/plano-plurianual-2016-2019-programacao-qualitativa> | |
| **Público-alvo:**  Sem restrição de acesso aos dados. | **Tipo de Dados:** Dados referente ao Plano Plurianual do Governo Federal dos anos de 2016 a 2019 |
| **Licença:** Domínio Público  **Última atualização dos dados:** 28/11/2018 14:10:02  **Formato dos dados:** zip+csv | |

|  |
| --- |
| **Palavras chaves:**  Empreendimentos; Indicadores; Metas; Objetivos; PPA; Plano Plurianual; Plano de Governo; Programa temático |

**ÍNDICE**

[1. OBJETIVO 1](#_Toc160004919)

[2. CATÁLOGO DE DADOS 2](#_Toc160004923)

[3. DESENVOLVIMENTO 3](#_Toc160004923)

[4. REFERÊNCIAS, NORMAS e/ou DOCUMENTOS COMPLEMENTARES 4](#_Toc160004924)

[5. ANEXOS 5](#_Toc160004925)

1. OBJETIVO

Elaborar uma análise dos dados do governo federal relacionados ao plano plurianual dos anos de 2016 a 2019. Essas análises são destinadas a responder perguntas pré-selecionadas e descritas ao longo desse documento. Como forma complementar além das automatizações realizadas em Python, foi construído um dashboard em Power BI para a melhor visualização e respostas dos dados.

2. CATÁLOGO DE DADOS

O catálogo de dados foi construído visando simplificar o processo de identificação dos dados, sendo assim através do catálogo é possível descobrir rapidamente quais são os datasets disponíveis no contexto como um todo, quais são as colunas, valores mínimos e máximos, tipo de dados e qual a finalidade de cada conjunto de dados disponível.

Antes da exibição de cada catálogo de dados é interessante ressaltar que existem tabelas com os mesmos tipos de dados, se diferenciando apenas nas temporalidades dos dados.

Interface gráfica do usuário, Tabela

Descrição gerada automaticamente

Figura 1: Base de dados contidas na fonte

Texto

Descrição gerada automaticamente

Figura 2: Abertura dos dados dos datasets (código) - “iniciativas1”

Texto

Descrição gerada automaticamente

Figura 3: Abertura dos dados dos datasets (código) - “ind\_ref”

Texto

Descrição gerada automaticamente com confiança média

Figura 4: Abertura dos dados dos datasets (código) - “iniciativas2”

Interface gráfica do usuário, Texto, Aplicativo

Descrição gerada automaticamente

Figura 5: Abertura dos dados dos datasets (código) - “programas”

Texto

Descrição gerada automaticamente

Figura 6: Abertura dos dados dos datasets (código) - “orgao1”

Texto

Descrição gerada automaticamente

Figura 7: Abertura dos dados dos datasets (código) - “orgao2”

Interface gráfica do usuário, Texto, Aplicativo, Email

Descrição gerada automaticamente

Interface gráfica do usuário

Descrição gerada automaticamente com confiança média

Interface gráfica do usuário, Texto

Descrição gerada automaticamente com confiança média

Interface gráfica do usuário, Texto, Aplicativo

Descrição gerada automaticamente

Figura 8: Abertura dos dados dos datasets (código) - “completa”

1. DESENVOLVIMENTO

Esse documento foi elaborado com base nos dados disponibilizados através do portal de dados do governo brasileiro (dados.gov.br), e seguiu os seguintes passos:

1. Busca dos dados
2. Coleta
3. Modelagem
4. Carga
5. Análise

Ao final desses passos pretende-se responder as seguintes perguntas:

1. Qual total de programas e iniciativas existentes em cada região?
2. Qual a média de custo por cada Órgão existente?
3. Qual a média de custo por exercício?
4. Relação da média dos valores gastos por região e Órgão
5. Total de Programas e iniciativas por exercício
6. Relação de quantidade de programas e iniciativas por exercício e Órgão
7. Quais são os programas existentes?
8. Ranking de programas e iniciativas
9. Média de iniciativas por programa existente

Após conhecer toda a estrutura dos dados e definir as perguntas que o projeto se propões a responder, o código em python ficou estruturado da seguinte forma:

1. Importação de bibliotecas para:
   1. Interface do usuário: Tkinter.
   2. Requisição a sites: Requests.
   3. Movimentação de documentos: Pathlib, os.
   4. Tratamento de base de dados: csv, pandas, zipfile, io.
2. Interface com o usuário para seleção do local onde será armazenado os dados brutos, ou seja, sem tratamento.

Texto

Descrição gerada automaticamente

1. Criação de funções para listar os arquivos zip e carregar os dataframes:

Texto

Descrição gerada automaticamente

1. Realizar a requisição a fonte de dados, listar os arquivos csv dentro do zip, salvar os arquivos com os dados brutos no local especificado.

Texto

Descrição gerada automaticamente

Interface gráfica do usuário, Texto, Aplicativo, chat ou mensagem de texto

Descrição gerada automaticamente

1. Realizando a primeira limpeza dos dados

Texto

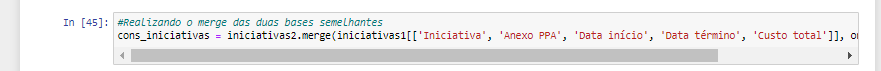
Descrição gerada automaticamente

1. Concatenando os dataframes com o mesmo conjunto de dados

Uma imagem contendo Calendário

Descrição gerada automaticamente

1. Realizar merge dos dataframes iniciativas1 e iniciativas2 visando terem o mesmo conjunto de dados para que tenha possibilidade de junção.



1. Removendo duplicatas, convertendo os padrões de datas, valores e retornando a quantidade de iniciativas

Texto

Descrição gerada automaticamente

1. Trabalhando os programas com o mesmo objetivo das iniciativas, mesclagem de dados, remoção de duplicatas, renomeação de colunas etc.

Uma imagem contendo Calendário

Descrição gerada automaticamente

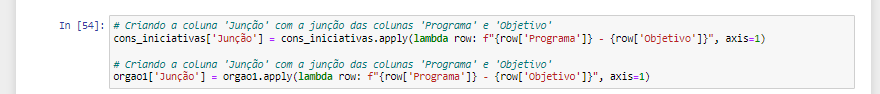
Texto

Descrição gerada automaticamente

Texto

Descrição gerada automaticamente

1. Juntando em um só dataframe consolidado, salvando e exibindo mensagem para o usuário.



Uma imagem contendo Texto

Descrição gerada automaticamente

Interface gráfica do usuário, Texto, Aplicativo, Email

Descrição gerada automaticamente

Finalizando a estrutura de código, para auxiliar foi realizada a criação de um Power BI para ajuda na visualização dos dados e respostas das perguntas. Abaixo link e print do Power BI.

<https://app.powerbi.com/reportEmbed?reportId=dac3f351-1399-48da-8248-79ede54c383b&autoAuth=true&ctid=0c6c23de-546b-45ff-811a-a88cc514ae5f>

Interface gráfica do usuário, Aplicativo, Tabela, Excel

Descrição gerada automaticamente

1. REFERÊNCIAS, NORMAS e/ou DOCUMENTOS COMPLEMENTARES

N/A

1. ANEXOS



Power BI – Plano Plurianual 2016-2019 – Anexo 1



Arquivo contendo o código em Python – Anexo 2