



FACULDADE DE TECNOLOGIA E INOVAÇÃO
FACSENAC DF

CÉLIDA SODRÉ
DÉBORA FERREIRA
LUDMILLA CARVALHO

Viagens a Serviço: 2019-2022

Análise Preditiva

Brasília, DF

2024

ÍNDICE

1. Introdução
2. Objetivo
3. Rótulos
4. Tipos de Dados
5. Quantitativos
6. Número de Datasets
7. Relacionamentos
8. Formato dos Dados
9. Análise Descritiva
10. Análise Diagnóstica
11. Análise Preditiva
12. Análise Prescritiva

1. INTRODUÇÃO

Neste documento, vamos explorar conjuntos de dados disponíveis no site portaldatransparência.gov, realizando uma análise preditiva com os conjuntos dos anos de 2019, 2020, 2021 e 2022, com o propósito de prever gastos futuros. O objetivo central é analisar, compreender e preparar os dados, utilizando seis métodos essenciais.

Rótulos: Identificaremos e descreveremos os rótulos, que representam as variáveis no dataset.

Tipos de Dados: Examinaremos os tipos de dados, diferenciando entre valores numéricos, textuais e datas, para entender como as informações estão formatadas.

Qualitativos: Analisaremos atributos qualitativos, que se referem a dados expressos em números e representam quantidades mensuráveis.

Número de Datasets: Investigaremos quantos conjuntos de dados estão incluídos na documentação.

Relacionamento: Descreveremos como os dados em diferentes datasets estão conectados.

Formatos dos Dados: Descreveremos os formatos em que os dados estão apresentados, como CSV ou Excel.

Além disso, realizamos a etapa de limpeza de dados no ambiente Colab para eliminar entradas duplicadas, tratamento de NaN eliminando a redundância e garantindo que as análises subsequentes sejam baseadas em dados confiáveis.

2. OBJETIVO

Nosso objetivo é realizar uma análise preditiva sobre os gastos futuros dos órgãos, identificando tendências de despesas e picos em períodos específicos, com uma precisão de 70%. Como estudantes de Ciência de Dados, estamos empenhados em obter insights valiosos sobre o futuro dos gastos públicos e seu impacto nas políticas e na sociedade. Compreender como os gastos evoluirão é essencial para tomar decisões bem fundamentadas sobre políticas públicas e alocação de recursos.

Durante o processo, realizaremos uma minuciosa coleta e análise de informações sobre os diversos setores de gastos, visando identificar padrões e oportunidades de otimização. Este escopo permitirá uma compreensão mais profunda dos fluxos financeiros e uma identificação clara de áreas onde a alocação de recursos públicos pode ser mais eficiente.

Para alcançar nossos objetivos, faremos uso de técnicas avançadas de análise preditiva, como modelagem estatística e aprendizado de máquina. Essas ferramentas nos permitirão extrair padrões complexos dos dados e criar modelos que possam prever com precisão os gastos futuros, fornecendo insights valiosos para embasar decisões estratégicas e políticas.

3. RÓTULOS

IDENTIFICADOR DO PROCESSO DE VIAGEM
NÚMERO DA PROPOSTA (PCDP)
SITUAÇÃO
VIAGEM URGENTE
JUSTIFICATIVA URGÊNCIA VIAGEM
CÓDIGO DO ÓRGÃO SUPERIOR
NOME DO ÓRGÃO SUPERIOR
CÓDIGO DO ÓRGÃO SOLICITANTE
NOME DO ÓRGÃO SOLICITANTE
CPF VIAJANTE
NOME
CARGO
FUNÇÃO
DESCRIÇÃO FUNÇÃO
PERÍODO – DATA DE INÍCIO
PERÍODO – DATA DE FIM
DESTINOS
MOTIVO
VALOR DIÁRIAS
VALOR PASSAGENS
VALOR DEVOLUÇÃO
VALOR OUTROS GASTOS

4 .TIPOS DE DADOS

Nome Coluna	Tipo de Dado	Descrição
Identificador do processo de viagem	Integer	Número que identifica o processo de concessão da viagem .
Número da proposta	Integer	Número que identifica a proposta da viagem
Código do Órgão Superior	Integer	Código do Órgão Superior que custeou a despesa ÓRGÃO SUPERIOR - Unidade da Administração Direta que tenha entidades por ele supervisionadas.
Nome Órgão Superior	Varchar	Nome do Órgão Superior
Código Órgão Pagador	Integer	Código do Órgão que pagou a despesa ÓRGÃO SUBORDINADO - Entidade supervisionada por um Órgão da Administração Direta.
Nome Órgão Pagador	Varchar	Nome do Órgão
Código UG Pagadora	Integer	Código da Unidade Gestora que pagou a despesa UNIDADE GESTORA (UG) - Unidade Orçamentária ou Administrativa que realiza atos de gestão orçamentária, financeira e/ou patrimonial

Nome UG Pagadora	Varchar	Nome da unidade gestora
Tipo de Pagamento	Varchar	Tipo da despesa paga pelo órgão (diária, passagem, seguro, etc.)
Valor da Despesa	Float	Valor da despesa paga.
Situação	Varchar	Situação da viagem: 'não' realizada; 'realizada'
Viagem urgente	Char	Viagem urgente: 'Sim', ou 'Não'
Justificativa	Varchar	Justificativa da urgência da viagem
Código solicitante	Integer	Código do órgão que solicitou a viagem
Nome do órgão solicitante	Varchar	Nome do órgão
CPF viajante	Integer	CPF do viajante
Nome	Varchar	Nome do viajante
Cargo	Varchar	Cargo do viajante
Função	Varchar	Registro da Função
Descrição	Varchar	Descrição da função
Período – Data de Início	Date	Data de início da viagem
Período – Data de Fim	Date	Data de fim da viagem
Destinos	Varchar	Destino da viagem
Motivos	Varchar	Motivo da viagem
Valor Diárias	Float	Valor das diárias
Valor Passagens	Float	Valor das passagens
Valor Devolução	Float	Valor devolução
Valor Outros Gastos	Float	Valor com outros gastos

5. QUANTITATIVOS

O conjunto de dados consiste em 2.266.556 registros distribuídos em 22 colunas.

6. NÚMERO DE DATASETS

Nesta pesquisa e análise, utilizamos apenas quatro conjuntos de dados.

7. RELACIONAMENTOS

Após a concatenação, ficamos apenas com um conjunto de dados.

8. FORMATO DE DADOS

Neste projeto, lidamos com arquivos nos formatos CSV.

9. Análise Descritiva

A análise descritiva dos dados de viagens a serviço revelou insights significativos sobre os padrões observados. Iniciamos pela exploração da estrutura do dataset, examinando e corrigindo valores nulos. Identificamos que destinos como São Paulo, Rio de Janeiro, Porto Alegre e Salvador se destacam como os mais frequentes em viagens corporativas pelo Brasil. Além disso, analisamos detalhadamente as viagens classificadas como urgentes e os cargos que mais frequentemente realizam essas viagens. Neste conjunto de dados:

- **Visualização do DataFrame:** Exibimos as primeiras linhas do dataset para compreender sua estrutura e os tipos de dados presentes.
- **Tratamento de Valores Nulos:** Realizamos a identificação e o tratamento de valores ausentes, removendo registros incompletos para assegurar a integridade dos dados analisados.
- **Informações do DataFrame:** Apresentamos uma visão geral das colunas, seus tipos de dados e a quantidade total de registros, fornecendo insights sobre a composição do dataset e facilitando análises subsequentes.

10. Análise Diagnóstica

A análise diagnóstica enriquece a abordagem descritiva ao explorar correlações e interações entre variáveis, oferecendo insights profundos sobre os padrões identificados nos dados de viagens corporativas. Ao investigar os destinos mais frequentes e os cargos com maior incidência de viagens urgentes, pudemos identificar áreas críticas que demandam atenção especial na gestão dessas viagens. Neste conjunto de dados específico:

- **Distribuição das Viagens Urgentes e Não Urgentes:** Analisamos a distribuição total entre viagens urgentes e não urgentes, destacando a prevalência das primeiras.
- **Top 5 Cargos com Viagens Urgentes:** Identificamos os cinco cargos que mais realizaram viagens urgentes, fornecendo insights sobre os perfis ocupacionais mais afetados pela necessidade de deslocamento imediato.
- **Distribuição de Urgência nos 5 Principais Destinos:** Investigamos a distribuição da urgência em relação aos cinco destinos mais frequentemente visitados, permitindo uma compreensão detalhada das demandas específicas de cada localidade.

11. Análise Preditiva

A análise preditiva utiliza modelos de machine learning para antecipar comportamentos futuros com base em dados históricos. Neste estudo, desenvolvemos um modelo Random Forest para classificar a urgência das viagens com base em variáveis relevantes. Avaliamos a acurácia do modelo para garantir sua eficácia na previsão dos padrões de viagem. Este passo é essencial para prever necessidades e tomar decisões estratégicas com antecedência. Neste conjunto de dados específico:

- **Codificação de Variáveis Categóricas:** Utilizamos codificação de rótulos para transformar variáveis categóricas em numéricas, permitindo que os modelos de machine learning processem essas informações.
- **Divisão dos Dados:** Separamos os dados em conjuntos de treinamento e teste para avaliar o desempenho do modelo em dados não vistos.
- **Padronização dos Dados:** Aplicamos o StandardScaler para padronizar as features numéricas do conjunto de dados, garantindo que todas as variáveis tenham a mesma escala.
- **Criação e Treinamento do Modelo Random Forest:** Desenvolvemos um modelo de classificação Random Forest e o treinamos com os dados de treinamento, ajustando-o para capturar os padrões nas viagens corporativas.
- **Avaliação do Modelo:** Calculamos a acurácia do modelo e geramos um relatório de classificação para entender sua capacidade de prever com precisão a urgência das viagens corporativas.

12. Análise Prescritiva

A análise prescritiva converte insights em ações práticas e estratégicas. Com base nos resultados das análises anteriores de nosso dataset de viagens corporativas, recomendamos medidas para otimizar a gestão de viagens urgentes e melhorar a eficiência operacional da organização:

- **Revisão de Políticas de Viagem:** Estabelecer critérios claros para identificação e tratamento de viagens urgentes, incluindo definição de limites de gastos e procedimentos de aprovação específicos.
- **Implementação de Tecnologias de Monitoramento:** Adotar sistemas de monitoramento em tempo real para identificar viagens urgentes potenciais. Integração de alertas automáticos para gestores sobre viagens que requerem atenção imediata.
- **Incentivo à Flexibilidade:** Promover o uso de ferramentas de colaboração online e videoconferências como alternativas viáveis para reuniões presenciais, reduzindo a necessidade de deslocamentos.
- **Relatórios e Transparência:** Desenvolver relatórios periódicos detalhando o número de viagens urgentes, seus destinos e os custos associados. Implementar painéis de controle para uma visualização transparente das despesas relacionadas a viagens urgentes.

Essas iniciativas não apenas visam reduzir custos operacionais, mas também promover maior transparência e controle na gestão de viagens urgentes, alinhando-se aos objetivos estratégicos da organização.