

UNIVERSIDADE PRESBITERIANA MACKENZIE

GUSTAVO JOSÉ FERMIANO – RA: 10440929

**PROJETO APLICADO I - ANÁLISE DOS REGISTROS PÚBLICOS DE
PAGAMENTOS DE DIÁRIAS E PASSAGENS EM VIAGENS REALIZADAS A
SERVIÇO, DENTRO DA ADMINISTRAÇÃO PÚBLICA FEDERAL, EM 2023.**

Campinas
2024

SUMÁRIO

Lista de figuras	3
Lista de tabelas.....	4
Introdução.....	5
Proposta Analítica.....	6
Análise Exploratória de Dados.....	11
Glossário.....	25
Referências Bibliográficas.....	26
Anexo I – Esboço do Storytelling.....	27
Anexo II – Vídeo de apresentação.....	30

LISTA DE FIGURAS

Figura 1 - Gráfico de viagens realizadas e não realizadas

Figura 2 - Gráfico de viagens de acordo com situação de urgência

Figura 3 - Gráfico de frequência de diárias por mês de início, no ano de 2023

Figura 4 - Ano de encerramento de viagens iniciadas em 2023

Figura 5 - Quantidade de diárias emitidas por dias de afastamento para viagem a serviço

Figura 6 - Frequência de viagens protegidas por sigilo

Figura 7 - Curva de densidade do valor das diárias

Figura 8 - Curva de densidade do valor das passagens

LISTA DE TABELAS

Tabela 1 – Descrição dos atributos do dataset de viagens a serviço

Tabela 2 – Número de diárias emitidas por órgão superior

Tabela 3 – Número de diárias emitidas por órgão subordinado

Tabela 4 – Cargos com maior frequência, considerando apenas as diárias registradas com órgão “Sem informação”

Tabela 5 – Destinos com maior número de emissão de diárias, no ano de 2023

Tabela 6 – Número de diárias por faixa de valor

Tabela 7 – Número de diárias por faixa do valor gasto em passagens

Tabela 8 – Número de diárias por faixa de valor devolvido

Tabela 9 – Número de diárias por faixa de valor dispendido em “outros gastos”

Tabela 10 – Matriz de correlação entre o total de dias de viagem e os valores gastos com diárias, passagens, valores devolvidos e outros gastos

Tabela 11 – Resumo das Medidas de Tendência Central e Dispersão

INTRODUÇÃO

Anualmente, são gastos mais de R\$ 1 bilhão de reais em pagamentos de diárias e passagens para atender servidores do quadro das instituições da Administração Direta e Indireta do Governo Federal.

As diárias são valores pagos a servidores públicos para cobrir despesas relacionadas à alimentação, hospedagem e outras despesas de pequeno vulto, quando os servidores precisam atuar fora de suas sedes de trabalho. Regulamentadas por lei, essas diárias variam de acordo com o cargo, o destino, e principalmente, ao número de dias do afastamento.

Já as passagens visam a cobrir os custos de transporte dos servidores durante estas viagens a serviço. Esses gastos podem ser intermediados por outras empresas, e aplicam-se a uma variedade de modais de transporte, em especial o modal rodoviário e as passagens aéreas.

Essas despesas, destinadas a custear viagens a serviço, passam por diversas instâncias de controladoria e fiscalização, que visam a evitar o mau uso dos recursos públicos. Devido ao grande volume de viagens a serviços realizadas todos os anos, demandam-se análises mais sofisticadas para detectar pagamentos irregulares ou indevidos.

Este trabalho se propõe a realizar a Análise Exploratória de Dados do *dataset* de Diárias e Viagens a Serviço do ano de 2023, disponibilizado pelo Governo Federal através do Portal da Transparência, aprofundando-se e produzindo informações sobre a distribuição desses pagamentos ao longo deste exercício, pela perspectiva da estatística descritiva.

PROPOSTA ANALÍTICA

O *dataset* utilizado está disponibilizado foi a base de dados de viagens a serviço do ano de 2023, disponibilizado no site do Portal da Transparência do Governo Federal: <https://portaldatransparencia.gov.br>. Este portal integra e apresenta dados de diversos sistemas utilizados pelo Governo Federal para a sua gestão financeira e administrativa. A frequência de atualização desses dados é diária, semanal ou mensal, a depender de cada tema investigado.

Este *dataset*, que compila as informações de passagens e diárias e serviços de todos os ministérios da administração pública federal, depende dos órgãos fornecidos pelos diferentes ministérios e órgãos da administração pública federal, e é mantido pela Controladoria Geral da União (CGU).

O site institucional da CGU define o espectro de atuação da entidade:

“A Controladoria-Geral da União (CGU) é o órgão do governo federal responsável pela defesa do patrimônio público e pelo incremento da transparência na gestão, por meio de ações de controle interno, auditoria pública, correição, ouvidoria e prevenção e combate à corrupção.

A CGU também exerce, como órgão central, a supervisão técnica dos órgãos que compõem o Sistema de Controle Interno, o Sistema de Correição (Siscor), o Sistema de Ouvidoria (SisOuv) e o Sistema de Integridade Pública (Sipef) do Poder Executivo federal, prestando a orientação normativa necessária.”

A visão da CGU, definida em seu mapa estratégico (2020-2023), é: “Ser reconhecida pelo cidadão como indutora de uma Administração Pública íntegra, participativa, transparente, eficiente e eficaz.”

Sua missão institucional é: “Elevar a Credibilidade do Estado por meio da participação social, do controle interno governamental e do combate à corrupção em defesa da sociedade”

Para cumprir essa missão institucional, o órgão possui em 2024 3.066 servidores em exercício, além de 1.008 aposentados e 147 pensionistas.

Utilizar uma base de dados pública possui diversas vantagens, em especial quanto à boa estruturação do dataset, a gratuidade dos dados e a publicidade de sua divulgação, nos termos da Lei de Acesso à Informação (LAI), independentemente de solicitação.

A principal limitação relacionada aos dados se refere aos dados de viagens à serviço com sigilo vigente decretado por lei.

A lei nº 12.527, de 18 de novembro de 2011, conhecida como a Lei de Acesso à Informação (LAI), determina a publicidade dos dados como preceito geral e o sigilo como exceção (Art. 3º – inciso I), porém também elenca as condições para que uma informação seja tratada com sigilo, em seu Art. 23:

“Art. 23. São consideradas imprescindíveis à segurança da sociedade ou do Estado e, portanto, passíveis de classificação as informações cuja divulgação ou acesso irrestrito possam:

I - pôr em risco a defesa e a soberania nacionais ou a integridade do território nacional;

II - prejudicar ou pôr em risco a condução de negociações ou as relações internacionais do País, ou as que tenham sido fornecidas em caráter sigiloso por outros Estados e organismos internacionais;

III - pôr em risco a vida, a segurança ou a saúde da população;

IV - oferecer elevado risco à estabilidade financeira, econômica ou monetária do País;

V - prejudicar ou causar risco a planos ou operações estratégicas das Forças Armadas;

VI - prejudicar ou causar risco a projetos de pesquisa e desenvolvimento científico ou tecnológico, assim como a sistemas, bens, instalações ou áreas de interesse estratégico nacional;

VII - pôr em risco a segurança de instituições ou de altas autoridades nacionais ou estrangeiras e seus familiares; ou

VIII - comprometer atividades de inteligência, bem como de investigação ou fiscalização em andamento, relacionadas com a prevenção ou repressão de infrações”

Em resumo, o sigilo é imposto principalmente por questões de segurança, e possui impactos na divulgação desses dados por um período de até 25 anos. Considerando o subgrupo de dados que tem condição de sigilo dentro do *dataset*, a solução será tratar os dados sigilosos como ausentes ou faltantes, quando não for possível utilizá-los na análise.

O link direto de acesso à base de dados utilizada é: <https://portaldatransparencia.gov.br/download-de-dados/viagens/2023>. O arquivo se

encontra em formato *.CSV, possui origem aberta e pública, isto é, pode ser acessado por qualquer cidadão brasileiro, independentemente de solicitação. Sua sensibilidade é limitada pelos dados sigilosos, pois como já citado anteriormente, as leis federais, amparadas pela Constituição, determinam a publicidade como preceito geral, e o sigilo como exceção. Os dados provêm dos registros de diárias no Sistema de Concessão de Diárias e Passagens – SCDP e a execução financeira é realizada pelo Sistema Integrado de Administração Financeira do Governo Federal – SIAFI.

O *dataset* utilizado possui os seguintes atributos, descritos na tabela abaixo, apoiados com o uso do Dicionário de Dados, disponibilizado no Portal da Transparência.

Tabela 1 – Descrição dos atributos do dataset de viagens a serviço

Coluna	Descrição	Tipo de dado
Identificador do processo de viagem	Número que identifica o processo de concessão da viagem	Numérico
Número da Proposta (PCDP)	Número de proposta gerado no Sistema de Concessão de Diárias e Passagens (SCDP)	Caracteres
Situação	Situação da viagem: 'não realizada'; 'realizada'	Caracteres
Viagem Urgente	Se a viagem foi definida como urgente: 'sim' ou 'não'	Caracteres
Justificativa Urgência Viagem	Justificativa registrada caso a viagem tenha sido inserida como urgente	Caracteres

Coluna	Descrição	Tipo de dado
Código do órgão superior	Código do Órgão Superior (unidade da Administração Direta que tenha entidades por ele supervisionadas) que custeou despesas da viagem	Numérico
Nome do órgão superior	Nome do Órgão Superior	Caracteres
Código órgão solicitante	Código do Órgão (entidade supervisionada por um órgão da administração direta) que solicitou a viagem	Numérico
Nome órgão solicitante	Nome do Órgão	Caracteres
CPF viajante	CPF da pessoa que realizou a viagem	Caracteres (parte dos algarismos está anonimizada)
Nome	Nome do viajante	Caracteres
Cargo	Cargo do viajante	Caracteres
Função	Função gratificada ou de confiança (se houver)	Caracteres
Descrição Função	Descrição da função (se houver)	Caracteres
Período – Data de início	Data de início de afastamento do servidor	Caracteres (Data no formato dd/mm/aaaa)
Período – Data de fim	Data de fim de afastamento do servidor	Caracteres (Data no formato dd/mm/aaaa)
Destinos	Locais pelos quais o viajante passará durante a viagem	Caracteres

Coluna	Descrição	Tipo de dado
Motivo	Motivo da viagem	Caracteres
Valor diárias	Valor de diárias pagas pelo órgão, se houver	Numérico (moeda, com separação de centavos com vírgula)
Valor passagens	Valor de passagens pagas pelo órgão, se houver	Numérico (moeda, com separação de centavos com vírgula)
Valor devolução	Valor devolvido, se houver	Numérico (moeda, com separação de centavos com vírgula)
Valor outros gastos	Valor de outros gastos pagos pelo órgão, se houver	Numérico (moeda, com separação de centavos com vírgula)

Através da análise exploratória dos dados disponíveis no Portal da Transparência, o estudo tem como objetivo identificar padrões e insights dentro da base de dados de viagens a serviço do Governo Federal no ano de 2023. Conforme definido na etapa 1, as metas estabelecidas pelo estudo são:

- Identificar os órgãos superiores com maiores gastos anuais com diárias e passagens.
- Identificar a dispersão dos pagamentos (média, mediana, desvio padrão), mapeando a distribuição da aplicação desses recursos por faixas de pagamento
- Identificar destinos mais frequentes das viagens a serviço.
- Identificar a evolução destes pagamentos ao longo do ano de 2023, verificando fatores de sazonalidade ou contenção de recursos.

ANÁLISE EXPLORATÓRIA DOS DADOS

A etapa de Análise Exploratória de Dados foi realizada dentro do ambiente do Google Colab, usando a linguagem Python. Foram importadas as bibliotecas Pandas e Matplotlib. Os dados produzidos foram armazenados no Github no link: <https://github.com/gustavo-fermiano/portran>, incluindo os scripts utilizados na análise. Devido ao tamanho da base de dados disponível (acima de 400 mb), o arquivo foi subido no Google Drive e importado diretamente para a plataforma do Google Colab.

A base de dados disponível contém 827.433 entradas, e inicialmente contava com 22 atributos. Foram gerados atributos adicionais dentro do ambiente Python, a partir da combinação de alguns atributos originais do *dataset*.

Inicialmente, verificamos a porcentagem de viagens contidas em nossa base de dados que foram Realizadas ou Não Realizadas. Essa informação é importante para a produção de alguns indicadores, em que filtrarei apenas as viagens que foram de fato realizadas, pois para estas houve necessariamente o dispêndio de recursos para diárias e passagens.



Observa-se a partir deste gráfico que a grande maioria das viagens a serviço registradas na base de dados foram de fato realizadas. Analisamos também as viagens de acordo com a situação de urgência.



O sistema de lançamento de diárias considera uma viagem urgente como aquela que foi solicitada com antecedência inferior a 15 dias. Verificamos assim que, no geral, a maioria das diárias são solicitadas com antecedência inferior a 15 dias. Trata-se de um índice relativamente alto, pois se esperaria que a maioria das diárias fosse solicitada com planejamento maior pelas unidades.

Verifica-se também quais são os órgãos superiores que realizaram a maioria da solicitação de diárias no período. A informação produzida pode ser visualizada na Tabela 2, abaixo.

Tabela 2 – Número de diárias emitidas por órgão superior

Nome do órgão superior	Número de diárias
Ministério da Educação	172.211
<i>Sem informação</i>	163.710
Ministério da Defesa	141.907
Ministério da Justiça e Segurança Pública	110.707
Ministério do Meio Ambiente e Mudança do Clima	42.313
Ministério do Planejamento e Orçamento	33.787
Ministério da Agricultura e Pecuária	28.956
Ministério da Saúde	21.684
Ministério da Previdência Social	19.614
Ministério dos Transportes	18.065
Ministério do Desenvolvimento Agrário e Agricultura Familiar	14.417
Presidência da República	13.889
Ministério dos Povos Indígenas	9.134
Ministério das Relações Exteriores	5.718
Ministério da Integração e do Desenvolvimento Regional	4.407
Ministério de Minas e Energia	4.375
Ministério da Cultura	3.888
Advocacia-Geral da União	3.734
Ministério das Comunicações	3.460
Ministério dos Direitos Humanos e Cidadania	2.893
Ministério da Ciência, Tecnologia e Inovação	2.523
Ministério do Desenvolvimento, Indústria, Comércio e Serviços	2.470
Controladoria-Geral da União	2.087
Ministério da Fazenda	655
Ministério do Turismo	513
Ministério da Gestão e da Inovação em Serviços Públicos	233
Ministério do Trabalho e Emprego	183

Uma das informações que chama a atenção neste atributo é a quantidade de registros marcados como “Sem informação”. Esse número de dados faltantes se repete também ao verificar os órgãos subordinados. Nesse caso, como há um número bem superior de órgãos subordinados na base de dados (177), nos limitaremos apenas aos 20 com maior frequência.

Tabela 3 – Número de diárias emitidas por órgão subordinado

Nome do órgão subordinado	Número de diárias
<i>Sem informação</i>	163.710
Polícia Federal	69.204
Comando da Aeronáutica	67.829
Comando do Exército	52.774
Polícia Rodoviária Federal	41.215
Fundação Instituto Brasileiro de Geografia e Estatística	33.375
Ministério da Agricultura e Pecuária - Unidades com vínculo direto	28.956
Ministério da Defesa - Unidades com vínculo direto	21.223
Instituto Brasileiro do Meio Ambiente e dos Recursos Naturais Renováveis	20.966
Instituto Chico Mendes de Conservação da Biodiversidade	19.596
Instituto Nacional do Seguro Social	19.409
Ministério da Saúde - Unidades com vínculo direto	14.373
Presidência da República - Unidades com vínculo direto	13.489
Instituto Nacional de Colonização e Reforma Agrária	13.396
Fundação Nacional do Índio	9.134
Ministério da Educação - Unidades com vínculo direto	7.657
Departamento Nacional de Infraestrutura de Transportes	7.174
Ministério dos Transportes - Unidades com vínculo direto	5.776
Ministério das Relações Exteriores - Unidades com vínculo direto	5.637
Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira	4.736

Verificamos apenas as entradas que estavam como “Sem informação” para órgão superior e órgão subordinado, para verificar quais eram os cargos mais frequentes. As informações geradas (para as 20 entradas mais comuns), estão na tabela abaixo.

Tabela 4 – Cargos com maior frequência, considerando apenas as diárias registradas com órgão “Sem informação”

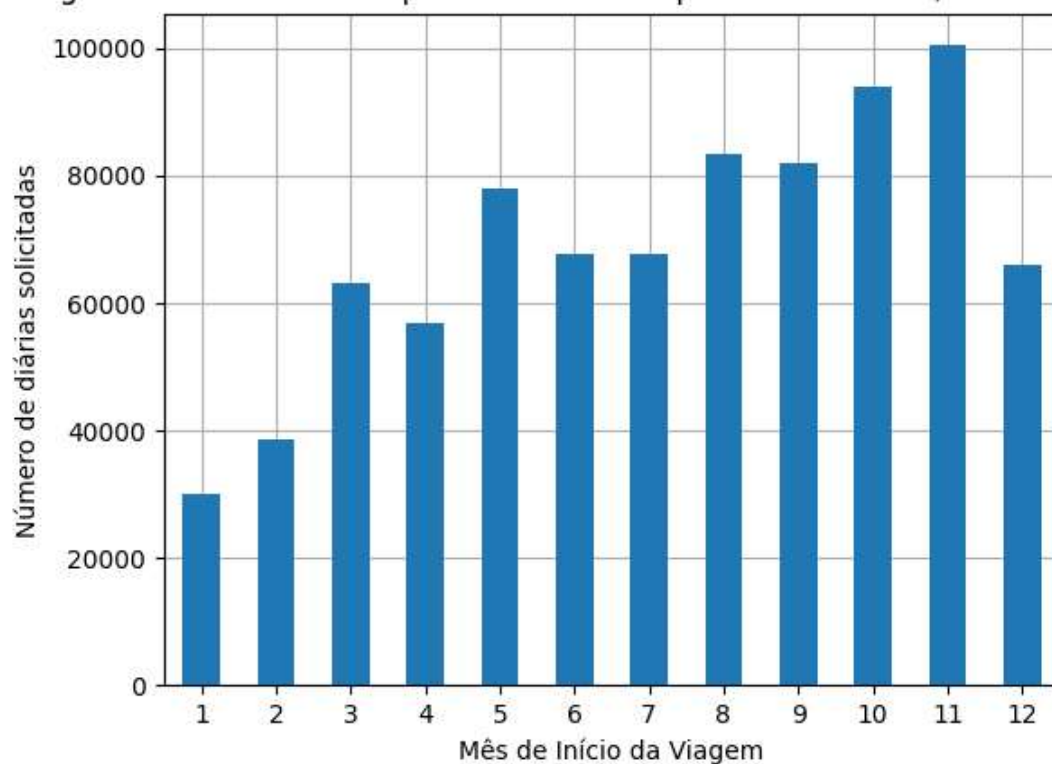
Cargo	Número de diárias
Auditor-Fiscal da Receita Federal	9.965
Analista Tributário da Receita Federal	8.505
Auditor Fiscal do Trabalho	7.760
<i>Informações protegidas por sigilo</i>	5.599
Agente Administrativo	2.365
Motorista Oficial	2.018
Professor do Magistério Superior	1.921
Assistente Técnico Administrativo	1.825
Especialista em Recursos Minerais	1.762
Analista de Infra-estrutura	1.083

Observa-se, dessa forma, que a maioria dos registros de diárias que constam como “Sem informação” referem-se a diárias emitidas para a Receita Federal, porém encontramos também incidências de cargos e funções vinculados a outros ministérios dentro da base de dados.

Foram analisadas também as diárias de acordo com a data de início e a data de fim das viagens a serviço. O objetivo era verificar quais eram os meses com maior incidência de pagamentos, e se haveria meses com represamento dos recursos orçamentários. O gráfico gerado ilustra bem como a distribuição dos recursos para pagamento de diárias e passagens em viagens a serviço obedece a princípios de sazonalidade, considerando as limitações orçamentárias impostas pelo próprio exercício financeiro.

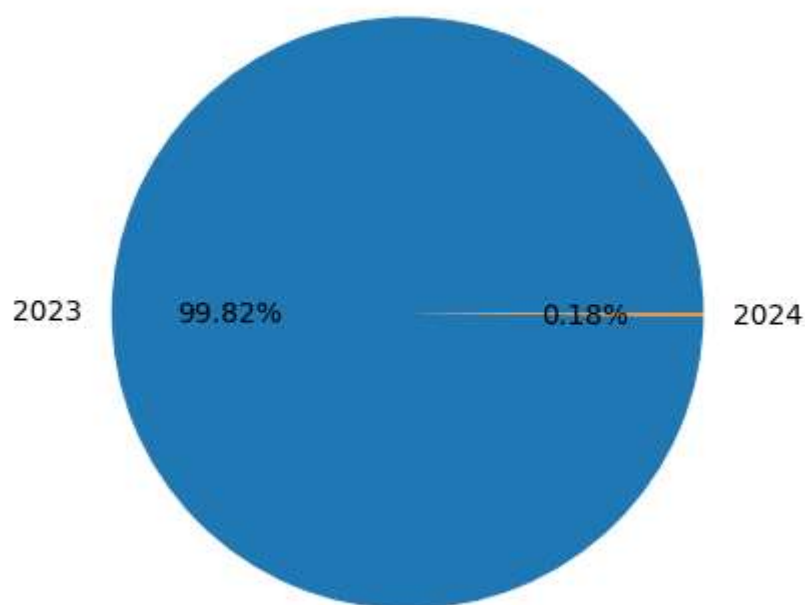
Observa-se que os meses de Janeiro, Fevereiro e Dezembro, devido ao fim e ao início do exercício financeiro, são caracterizados pelo represamento de recursos e redução do número de viagens a serviço. Os meses de Agosto a Novembro são caracterizados por uma concentração substancial das diárias.

Figura 3 - Gráfico de frequência de diárias por mês de início, no ano de 2023



Verifica-se também que todas as diárias do banco de dados disponível foram iniciados no próprio ano de 2023. Observa-se que a imensa maioria das diárias se inicia e encerra no próprio ano corrente, porém uma pequena quantidade se encerra no ano seguinte. A Figura 4 ilustra esse fato.

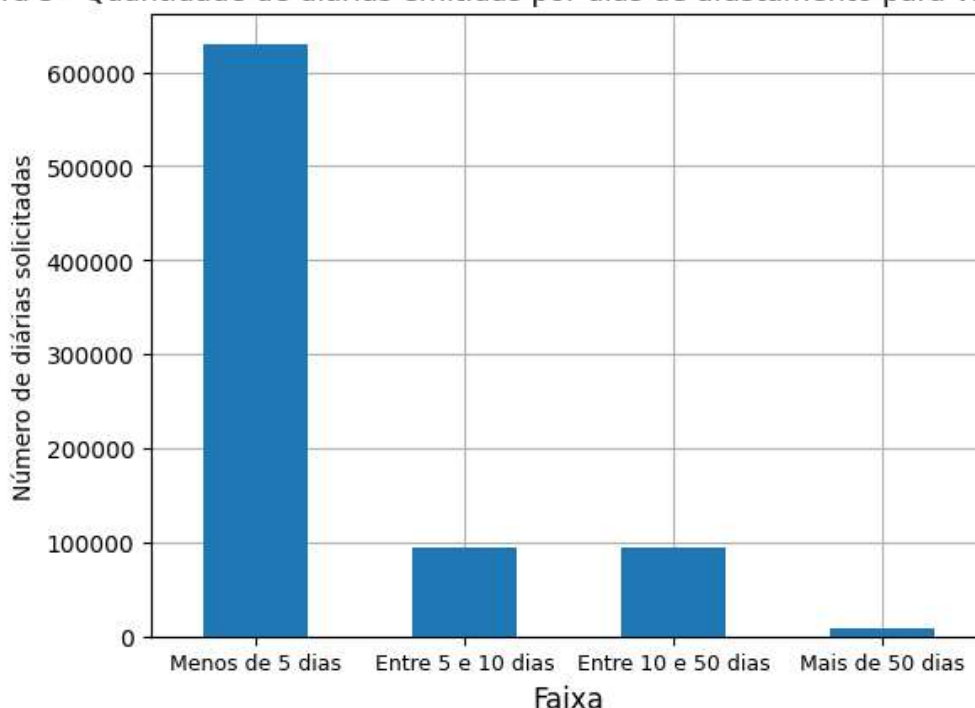
Figura 4 - Ano de encerramento de viagens iniciadas em 2023



Criei um novo atributo no *dataframe* com o número de dias total da viagem. A intenção foi, para além de visualizar os meses de início e fim das viagens a serviço, consultar informações sobre o período total das viagens a serviço.

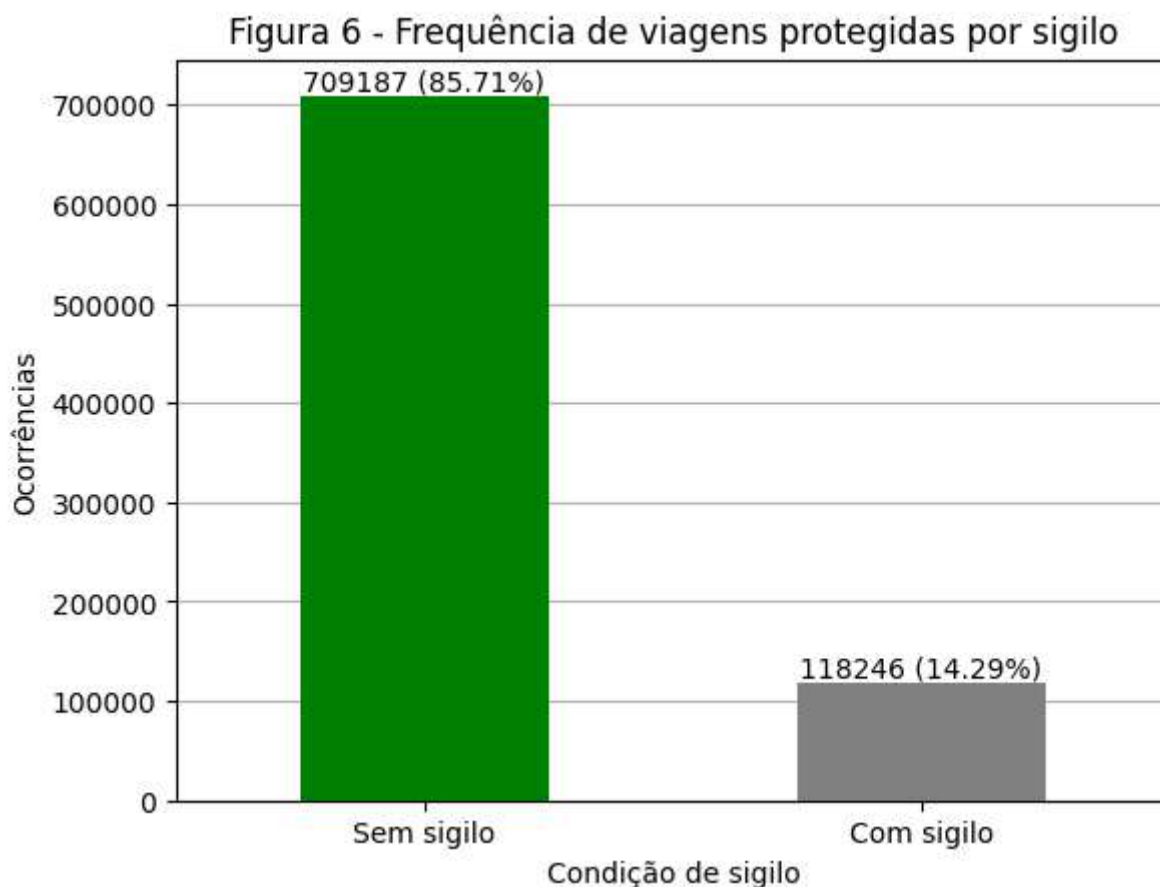
Observa-se que as viagens a serviço correspondem a uma média de 5,9 dias, com uma mediana de 3 dias de viagem. A diferença entre a média e a mediana nos sugere que há uma grande variação no número de diárias solicitado, e esse fato é confirmado pelo desvio padrão, de 18,07 dias. Para ilustrar melhor esse fato, foram criadas faixas para o tempo decorrido nas viagens a serviço: de 0 a 5 dias, de 5 a 10 dias, de 10 a 50 dias, e acima de 50 dias. Essa informação está graficamente ilustrada na figura 5.

Figura 5 - Quantidade de diárias emitidas por dias de afastamento para viagem a serviço



Observa-se, portanto, que embora a imensa maioria das diárias corresponda a períodos bem pequenos de afastamento (com mediana de 3 dias), uma quantidade pequena de servidores (aproximadamente 0,96%), que podemos chamar de “superviajantes”, se afasta por períodos superiores a 50 dias. De fato, encontramos uma amplitude de 525 dias, sendo o maior afastamento de 525 dias e o menor deles de apenas 0 dias, para o caso de viagens a serviço que iniciam e encerram no mesmo dia.

Verifiquei também o percentual de viagens sigilosas, conforme ilustrado na Figura 6. Essa informação é muito importante, devido ao fato de não podermos acessar a maioria das variáveis das viagens sigilosas. Observa-se que 14,29% das diárias emitidas no ano de 2023 tem suas informações protegidas por sigilo.



Analisei também os destinos mais frequentes de viagens a serviço no ano de 2023. Para esse cálculo, filtrei as viagens sigilosas. Os destinos com maior emissão de diárias estão representados na tabela 5.

Tabela 5 – Destinos com maior número de emissão de diárias, no ano de 2023

Destino	Número de diárias
Brasília/DF	64.320
Rio de Janeiro/RJ	26.254
São Paulo/SP	24.629
São José dos Campos/SP	12.289
Salvador/BA	9.009
Belo Horizonte/MG	8.690
Curitiba/PR	8.506
Recife/PE	8.211
Manaus/AM	7.407
Porto Alegre/RS	7.162
Belém/PA	6.896
Fortaleza/CE	6.710
Foz do Iguaçu/PR	6.549
Campo Grande/MS	6.214
Florianópolis/SC	5.926
Goiânia/GO	5.744
Natal/RN	5.225
João Pessoa/PB	4.703
Campinas/SP	4.703
Boa Vista/RR	4.632

Observa-se que os principais destinos viajados são nacionais, e que existe uma concentração de viagens destinadas a Brasília e o Rio de Janeiro, principalmente pela grande quantidade de sedes da administração pública estabelecidas nesses dois municípios.

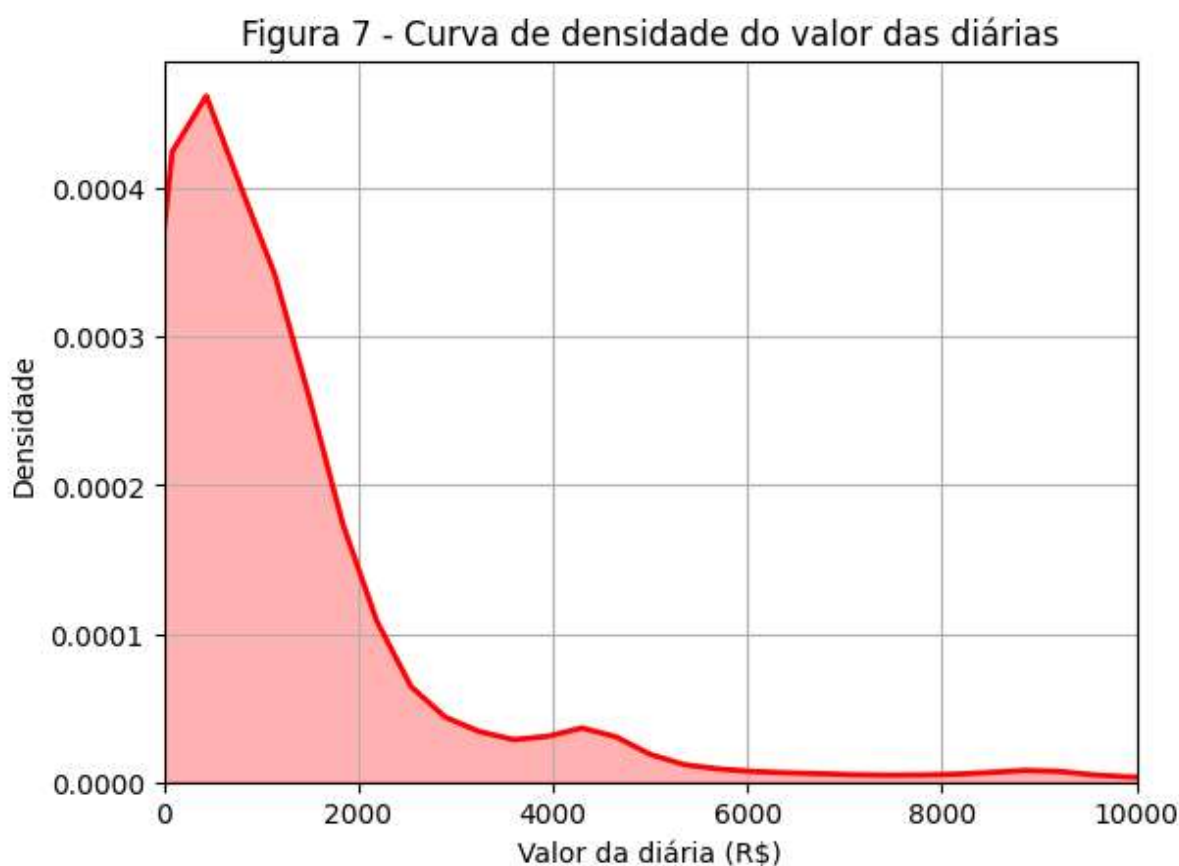
Depois disso analisa-se a coluna do valor pago de diárias.

Observa-se que o valor pago pelas diárias tiveram valor mínimo de R\$ 0,00 e valor máximo de R\$ 175.337,40. A variância foi de 24.264.315,60 (reais ao quadrado) e o desvio padrão de R\$ 4.925,88. Observa-se também, que não existem valores do tipo NA na coluna de pagamentos de diárias. A média foi de R\$ 1.744,25 e a mediana foi de R\$ 857,82.

O primeiro quartil (Q1) tem valor de R\$ 265,50, enquanto o terceiro quartil (Q3) tem valor de R\$ 1.629,59.

Ao tentar identificar a existência de outliers, verifiquei o limite inferior de R\$ 1.780,63 negativos, valor que foi desprezado, pois todos os pagamentos de diárias tiveram valores positivos. O limite superior para definição de outliers foi de R\$ 3.675,72, e foram identificados 80.116 outliers.

O entendimento é de que, devido a grande dispersão dos valores, conforme demonstrado nas outras variáveis, esses outliers representam corretamente valores pagos no exercício de 2023, e não devem ser desprezados. Isto porque a curva de densidade dos valores de diárias em nada se assemelha a uma curva normal, conforme ilustrado na figura 7.



A curva de densidade nos mostra uma curva de cauda longa (95% dos pagamentos são inferiores a R\$ 5.378,67, e 99% dos pagamentos são inferiores a R\$ 15.916,85), porém com alguns valores isolados acima de R\$ 100.000,00.

Podemos também ilustrar essa concentração de pagamentos por faixas de valor, conforme os dados na tabela 6.

Tabela 6 – Número de diárias por faixa de valor

Faixa de valor	Número de diárias	Percentual
Menos de R\$ 500,00	301.549	36,44%
Entre R\$ 500,00 e R\$ 1.000,00	157.846	19,08%
Entre R\$ 1.000,00 e R\$ 10.000,00	350.106	42,31%
Entre R\$ 10.000,00 e R\$ 100.000,00	17.589	2,13%
Acima de R\$ 100.000,00	343	0,04%

No atributo de valor das passagens, observa-se uma variância também bastante alta. Com um valor mínimo de R\$ 0,00 e um valor máximo de R\$ 186.723,29, o atributo tem variância de 6.523.803,64 (reais ao quadrado) e um desvio padrão de R\$ 2.554,17. Não existem valores NA no atributo. A média foi de R\$ 1.023,86 e a mediana foi zero.

Observa-se assim que ao menos metade das viagens não paga nenhum valor de passagens. O terceiro quartil é de R\$ 1.360,51, o percentil de 95% é de R\$ 4.934,31 e o percentil de 99% é de R\$ 9.597,33.

Observa-se um padrão semelhante ao dos valores de diárias, com um valor negativo no limite inferior e um limite superior modesto (R\$ 3.401,27), com a existência de 90.095 outliers na amostra.

A figura 8 ilustra a curva de densidade dos valores de passagens. Para ilustrar melhor, foi limitado o eixo de valor das passagens até o valor de R\$ 10.000,00. Observa-se a existência de um grande número de viagens que não paga nenhum valor ou muito próximo de zero, e outra tendência de cauda longa, com um pequeno pico na frequência ao redor do valor de R\$ 2.000,00.

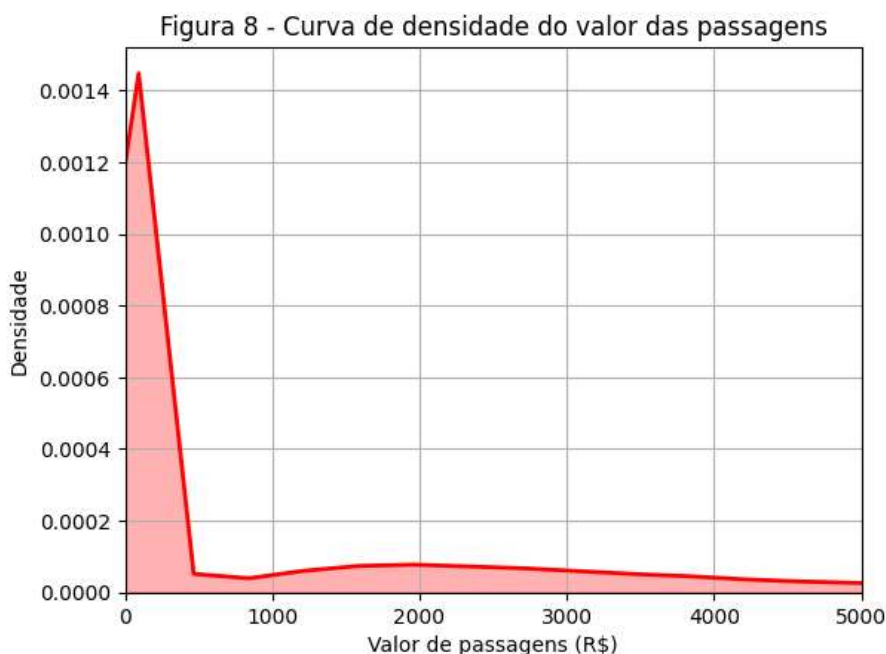


Tabela 7 – Número de diárias por faixa do valor gasto em passagens

Faixa de valor das passagens	Número de diárias	Percentual
Menos de R\$ 500,00	589.180	71,21%
Entre R\$ 500,00 e R\$ 1.000,00	13.805	1,67%
Entre R\$ 1.000,00 e R\$ 10.000,00	216.732	26,19%
Entre R\$ 10.000,00 e R\$ 100.000,00	7.699	0,93%
Acima de R\$ 100.000,00	17	Inferior a 0,01%

Em relação ao atributo de valor que foi devolvido pelos servidores (atributo “Valor devolução”) observa-se que o valor máximo foi de R\$ 38.884,71, e o valor mínimo de R\$ 0,00. Não existem valores NA no atributo. A média foi de apenas R\$ 23,74, e a mediana novamente foi zerada. A variância foi de 79.574,14 (reais ao quadrado) e o desvio padrão de R\$ 282,09.

Verificou-se que para 801.064 valores do dataset (ou seja, para 96,81% das diárias registradas) não houve nenhum valor a ser devolvido. A tabela 8 detalha melhor a distribuição dos valores de devolução por faixa.

Tabela 8 – Número de diárias por faixa de valor devolvido

Faixa de valor devolvido	Número de diárias	Percentual
Nenhum valor a ser devolvido	801.064	96,81%
Até R\$ 1.000,00	20.855	2,52%
Entre R\$ 1.000,00 e R\$ 10.000,00	5.404	0,65%
Entre R\$ 10.000,00 e R\$ 100.000,00	110	0,01%

Já em relação aos valores dispendidos com outros gastos dos servidores em viagens a serviço, o valor máximo registrado foi de R\$ 59.668,18, e o valor mínimo foi de R\$ 0,00. A média foi de apenas R\$ 16,40, e a mediana foi zerada. A variância foi de 23.170,61 (reais ao quadrado) e o desvio padrão de R\$ 152,22. Também não existem valores NA.

Verificou-se que para 790.437 valores do dataset (ou seja, para 95,53% das diárias registradas) não houve nenhum valor dispendido em “outros gastos”.

Tabela 9 – Número de diárias por faixa de valor dispendido em “outros gastos”

Faixa de valor dispendido em “outros gastos”	Número de diárias	Percentual
Nenhum valor com outros gastos	790.437	95,53%
Até R\$ 1.000,00	34.828	4,21%
Entre R\$ 1.000,00 e R\$ 10.000,00	2.164	0,26%
Entre R\$ 10.000,00 e R\$ 100.000,00	4	Menos de 0,01%

A partir das variáveis numéricas disponíveis, foi criada uma matriz de correlação conforme ilustrada na tabela 10.

Tabela 10 – Matriz de correlação entre o total de dias de viagem e os valores gastos com diárias, passagens, valores devolvidos e outros gastos

	Total de dias da viagem	Valor pago em diárias	Valor pago em passagens	Valor devolvido	Valor dispendido em outros gastos
Total de dias da viagem	1	0,89	0,05	0,10	0,02
Valor pago em diárias	0,89	1	0,15	0,14	0,10
Valor pago em passagens	0,05	0,15	1	0,07	0,19
Valor devolvido	0,10	0,14	0,07	1	0,05
Valor dispendido em outros gastos	0,02	0,10	0,19	0,05	1

A partir do coeficiente de Pearson, observamos que os elementos presentes que a maioria dos atributos possui, entre si, uma correlação próxima da nula ou mesmo muito baixa, porém positiva. A exceção, como já era de se prever, se trata

da correlação entre o valor pago em diárias e o total de dias da viagem em serviço. O valor de 0,89 sugere um forte grau de correlação linear positivo, sugerindo que de fato viagens mais longas demandam um dispêndio maior no valor do pagamento de diárias.

Por fim, a tabela 11 traz o resumo das medidas de tendência central e dispersão das principais variáveis numéricas do *dataframe*.

Tabela 11 - Resumo das Medidas de Tendência Central e Dispersão

	Valor Mín.	Valor Máx.	Variância	Desvio Padrão	Média	Mediana
Valor de diárias (R\$)	0,00	175.337,40	24.264.315,6	4.925,88	1.744,25	857,82
Valor de passagens (R\$)	0,00	186.723,29	6.523.803,64	2.554,17	1.023,86	0,00
Valor de devolução (R\$)	0,00	38.884,71	79.574,14	282,09	23,74	0,00
Valor de outros gastos (R\$)	0,00	59.668,18	23.170,61	152,22	16,39	0,00
Total de dias de viagem	0	525	326,47	18,07	5,90	3
Data de início da viagem	01/01/23	31/12/23	-	-	25/07/23	06/08/23
Data de fim da viagem	01/01/23	30/11/24	-	-	31/07/23	11/08/23

GLOSSÁRIO

Análise Exploratória de Dados (EDA) – processo de investigar e resumir um conjunto de dados para entender suas características principais, identificar padrões, detectar anomalias (outliers), testar hipóteses e verificar suposições.

Atributo – característica ou variável que descreve uma propriedade de um dado em um conjunto, representado por colunas em um *dataframe*.

Dataframe – estrutura de dados bidimensional, semelhante a uma tabela, que armazena e manipula dados em linhas e colunas.

Dataset – conjunto estruturado de dados, geralmente organizado em forma de tabela, usado para análise, modelagem ou treinamento de algoritmos.

Estatística descritiva – área da estatística que resume e descreve as características principais de um conjunto de dados, usando medidas como média, mediana e desvio padrão.

Exercício financeiro – período de 12 meses, durante o qual se realiza a gestão orçamentária e financeira de um governo ou empresa.

GitHub – plataforma online para hospedagem e gerenciamento de código-fonte, que facilita a colaboração e o controle de versão em projetos.

Google Colab – plataforma online que permite executar códigos em Python ou R em nuvem, com recursos gratuitos de computação.

NA – em ciência de dados, NA indica dados ausentes ou faltantes em um conjunto de dados.

Python – linguagem de programação de alto nível, usada para diversas aplicações como desenvolvimento web, análise de dados e inteligência artificial.

Script – conjunto de instruções de código que é executado sequencialmente para automatizar tarefas ou realizar ações em um programa.

REFERÊNCIAS BIBLIOGRÁFICAS

CONTROLADORIA GERAL DA UNIÃO – CGU. Portal da Transparência, 2024. Despesas do órgão: execução orçamentária e financeira. Disponível em: <<https://portaldatransparencia.gov.br/orgaos-superiores/37000-controladoria-geral-da-uniao>>. Acesso em: 29 de setembro de 2024.

GOVERNO FEDERAL. Gov.br. Controladoria Geral da União: Institucional. Disponível em: <<https://www.gov.br/cgu/pt-br/aceso-a-informacao/institucional>>. Acesso em: 29 de setembro de 2024.

GOVERNO FEDERAL. Gov.br. Mapa Estratégico: Controladoria-Geral da União 2020.2023. Disponível em: <<https://www.gov.br/cgu/pt-br/aceso-a-informacao/governanca/planejamento-estrategico/arquivos/mapa-estrategico-cgu.pdf>>. Acesso em: 29 de setembro de 2024.

CONTROLADORIA GERAL DA UNIÃO – CGU. Dicionário de Dados – Viagens a Serviço – Pagamentos. Disponível em: <<https://portaldatransparencia.gov.br/pagina-interna/603364-dicion%C3%A1rio-de-dados-viagens-a-Servi%C3%A7o-Pagamentos>>. Acesso em: 29 de setembro de 2024.

NEW CASTLE UNIVERSITY. Strength of Correlation. Disponível em: <<https://www.ncl.ac.uk/webtemplate/ask-assets/external/maths-resources/statistics/regression-and-correlation/strength-of-correlation.html>>. Acesso em: 29 de setembro de 2024.

ANEXO I – ESBOÇO DO STORYTELLING

Apresentação do grupo: Trabalho executado na modalidade individual.

Nome: Gustavo José Fermiano, RA: 10440929, E-mail: 10440929@mackenzista.com.br, aluno do 2º semestre do curso de Tecnologia em Ciência de Dados na Universidade Presbiteriana Mackenzie.

Nome do projeto: Projeto Aplicado I – Portran – Análise dos Registros Públicos de Pagamentos de Diárias e Passagens em Viagens Realizadas a Serviço, dentro da Administração Pública Federal, em 2023.

Empresa/Organização de estudo: Governo Federal – dados disponibilizados no Portal da Transparência pela Controladoria Geral da União (CGU).

Área do problema: A área do problema abrange a gestão dos recursos públicos federais, especificamente no que diz respeito ao pagamento de diárias e passagens para servidores em viagens a serviço. Anualmente, o Governo Federal gasta mais de R\$ 1 bilhão com esses pagamentos, envolvendo tanto despesas com alimentação e hospedagem quanto com transporte. A demanda crescente por viagens, a diversidade de órgãos envolvidos e a necessidade de controle sobre esses recursos tornam a análise desse processo crucial para garantir transparência e eficiência no uso dos recursos públicos.

Descrição do problema / gap: O problema identificado está na falta de uma análise mais aprofundada e estruturada dos dados relacionados aos pagamentos de diárias e passagens, especialmente quanto à identificação de possíveis irregularidades ou padrões inadequados de uso dos recursos públicos. Esses recursos precisam ser constantemente fiscalizados, e devido ao grande volume de pagamentos, é necessário estabelecer critérios mais claros para pagamentos suspeitos. O sigilo de parte das informações aprofunda ainda mais a dificuldade em fiscalizar adequadamente a destinação dos recursos públicos.

Proposta analítica: *item Proposta Analítica – página 6.*

Dados disponíveis: Base de dados de viagens a serviço do ano de 2023, disponibilizado no site do Portal da Transparência do Governo Federal: <https://portaldatransparencia.gov.br>. Este portal integra e apresenta dados de diversos sistemas utilizados pelo Governo Federal para a sua gestão financeira e administrativa. Essa base de dados compila as informações de passagens e diárias

e serviços de todos os ministérios da administração pública federal, depende dos órgãos fornecidos pelos diferentes ministérios e órgãos da administração pública federal, e é mantido pela Controladoria Geral da União (CGU).

Análise Exploratória: *item Análise Exploratória de Dados – página 11.*

Resultados pretendidos: Os resultados pretendidos foram a construção de recomendações para a fiscalização dos recursos de diárias e viagens a serviço, verificados na base de dados disponível.

Roteiro:

Segue planejamento de roteiro a seguir para elaboração de vídeo de 5 minutos de apresentação do Projeto Aplicado I.

1. Introdução - Apresentação do grupo e do nome do projeto – (0:30)

→ *Apresentar brevemente o nome, o curso, nome da universidade e o nome do projeto.*

2. Introdução - Contexto do projeto, instituição e área do problema – (0:30)

→ *Definição sucinta do escopo de atuação do projeto: a base de dados de diárias e passagens do Governo Federal no ano de 2023 e a importância da fiscalização dos recursos públicos aplicados, utilizando técnicas de estatística descritiva.*

3. Proposta Analítica – Apresentação da CGU e da base de dados (0:45)

→ *Explicação sucinta do escopo de atuação da Controladoria Geral da União e das principais características da base de dados utilizada (número de linhas, coluna, origem, publicidade, limitações, atributos categóricos e numéricos etc).*

4. Proposta Analítica – Apresentação dos objetivos iniciais (0:15)

→ *Apresentação dos objetivos iniciais apresentados na primeira etapa do projeto.*

5. Análise Exploratória de Dados – Descrição das ferramentas, biblioteca e linguagem utilizadas (0:15)

→ *Descrição sucinta das ferramentas, biblioteca e linguagem utilizadas na etapa de EDA.*

6. Análise Exploratória de Dados – Apresentação dos gráficos e tabelas produzidos na etapa de EDA (1:45)

→ *Apresentação sucinta dos principais gráficos e tabelas produzidos ao longo da etapa de EDA: elevado número de viagens urgentes, o Ministério da Educação como principal emissor de diárias, porém elevado número de órgãos sem informação, o levantamento de que a maior parte das diárias sem informação de órgão corresponde a servidores da Receita Federal, explicação do fator de sazonalidade (agosto a novembro) no pagamento de diárias, a predominância de viagens curtas, o percentual de viagens protegidas por sigilo, principais destinos. Salientando também a existência de grande número de outliers em pagamentos de diárias e viagens a serviço.*

7. Conclusão – Recomendações (0:45)

→ *São apresentadas as principais recomendações com base nos principais insights obtidos ao longo do projeto:*

- *Recomendação de adequação do nome do órgão para diárias emitidas com órgão “sem informação” – principalmente para servidores da Receita Federal*
- *Recomendação da revisão do critério de “urgência” para emissão de diárias*
- *Fiscalização de ‘super-viajantes’ e necessidade de melhor justificação para esses pagamentos: existência de uma curva bastante assimétrica com pagamentos muito acima da média*
- *Recomendação de equalizar o pagamento de diárias ao longo do ano, evitando concentração sazonal de diárias próximo ao fechamento do exercício financeiro*

8. Encerramento – Agradecimentos, apresentação de dados de contato e GitHub (0:15)

→ *Agradecimentos, apresentação de e-mail e link do GitHub do projeto.*

ANEXO II – VÍDEO DE APRESENTAÇÃO

O vídeo de apresentação foi gravado e está disponível no Youtube pelo link:

<https://www.youtube.com/watch?v=2LZvjY6Aer4>