

Documentação Rota02BI

Links:

- **Apresentação:** https://docs.google.com/presentation/d/1meomyQbB6i8WOrxNzL_c2PEQj3F1wkg-OJnBmYiHqYc/edit?usp=sharing
- **Dashboard:** <https://app.powerbi.com/view?r=eyJrJljoNDhhZDM3ZTEtYjZjYi00NGI3LWJmOTktMjVhNmRIMTdhNmYwliwidCl6ljhiZTBkOTY1LWE1NTktNDYyNC1iN>

Projeto de Exploração: Business Intelligence

1. Introdução

Graças à incessante coleta de dados que realizamos a todo momento, a tomada de decisões informada tornou-se um pilar fundamental para o sucesso empresarial. A Inteligência de Negócios (BI) emerge como uma ferramenta essencial, permitindo que organizações convertam grandes volumes de dados em informações cruciais para a tomada de decisões e ações. BI transcende a mera coleta de dados; envolve uma análise aprofundada e interpretação de padrões, oferecendo uma visão clara das operações comerciais e permitindo a identificação de oportunidades estratégicas, a melhoria da eficiência operacional e a manutenção da agilidade em um ambiente de negócios dinâmico e competitivo.

O que você aprenderá:

Neste projeto você aprenderá como relacionar tabelas no Power BI, a trabalhar com colunas calculadas e fórmulas DAX, além de criar um painel usando visualizações avançadas, como mapas geográficos. Essas habilidades te ajudaram em:

- Análise de dados complexos: relacionar tabelas no Power BI permite juntar conjuntos de dados de várias fontes. Isto é essencial quando você trabalha com dados complexos e precisa extrair informações valiosas de múltiplas fontes.
- Precisão de dados aprimorada: usando colunas calculadas e fórmulas DAX, você pode realizar cálculos personalizados e manipular os dados de acordo com as necessidades específicas de sua análise. Esse processo contribui para a precisão e relevância das informações que você apresenta.
- Tomada de decisão informada: a capacidade de criar painéis com visualizações avançadas, como mapas geográficos, fornecem uma maneira eficaz de representar dados visualmente. Isto facilita a identificação de padrões, tendências e relacionamentos, que por sua vez melhora a tomada de decisão informada.

Caso

Com o avanço das novas tecnologias, novos modelos de negócio também surgem, um exemplo claro disso são as plataformas como o Airbnb, que revolucionaram a maneira como as pessoas buscam e oferecem hospedagem. Nessa perspectiva empresarial, maximizar a eficiência e a rentabilidade tornou-se fundamental tanto para os anfitriões quanto para a própria plataforma.

Nesse contexto, este projeto concentra-se na exploração e análise de dados relacionados à disponibilidade de quartos no Airbnb, utilizando ferramentas e conceitos de negócios como Business Intelligence (BI) para revelar padrões, identificar oportunidades e aprimorar a tomada de decisões estratégicas.

A diversidade de informações geradas pela interação dos anfitriões e hóspedes na plataforma Airbnb resulta em um vasto conjunto de dados. Esses dados abrangem desde detalhes sobre imóveis, preços e localizações até feedback dos hóspedes. A riqueza desses dados oferece uma oportunidade única para aplicar técnicas de BI, como a integração de dados e a análise de relacionamentos entre tabelas. Buscamos descobrir insights valiosos que possibilitem aos anfitriões e ao próprio Airbnb otimizar a disponibilidade de quartos, maximizar a receita e aprimorar a experiência dos hóspedes.

Neste projeto, você realizará uma análise exploratória e utilizará técnicas avançadas de BI para visualizar tendências, identificar padrões e compreender os fatores que influenciam a ocupação do alojamento. Desde a sazonalidade elevada até as preferências regionais, diversos aspectos serão examinados para desvendar informações valiosas. O objetivo final é fornecer uma base sólida para a tomada de decisões informadas, permitindo que as partes interessadas adotem medidas estratégicas para aprimorar a eficiência operacional e lucratividade no dinâmico ecossistema do Airbnb.

1. Ferramentas, linguagens e insumos

1.1 Ferramentas e/ou plataformas

Neste projeto você usará duas ferramentas, uma para gerenciar os dados e outra para fazer a análise exploratória:

- Para gerenciar as tabelas, você usará o Google BigQuery
- Para realizar a análise exploratória dos dados você utilizará o PowerBI.

1.2 Linguagens

Você usará a linguagem SQL no BigQuery e as fórmulas DAX no PowerBI.

1.3 Insumos

Neste projeto você terá um conjunto de dados estruturados com 3 tabelas (2 tabelas de dimensão e 1 de transação, leia este artigo para entender melhor este [conceito](#) com informações sobre os quartos disponíveis para reserva no Airbnb).

O conjunto de dados está disponível para download neste [link](#).

Abaixo, você pode consultar a descrição das variáveis que compõem as tabelas:

Tabela rooms (Dimensão):

- **id**: um identificador exclusivo para cada quarto.
- **nome**: o nome do anúncio do Airbnb.
- **bairro**: sigla para o bairro em que o anúncio do Airbnb está localizado.
- **neighbourhood_group**: bairro onde o anúncio do Airbnb está localizado.
- **latitude**: a coordenada de latitude do anúncio do Airbnb.
- **longitude**: a coordenada de longitude do anúncio do Airbnb.
- **room_type**: o tipo de quarto oferecido no anúncio do Airbnb.
- **minimum_nights**: o número mínimo de noites aceito na reserva.

Tabela hosts (Dimensão):

- **host_id**: um identificador exclusivo para cada host.
- **nomedohost**: o nome do anfitrião do anúncio do Airbnb.

Tabela reviews (fatos):

- **id**: um identificador exclusivo para cada quarto.
- **host_id**: um identificador exclusivo para cada host.
- **preço**: o preço por noite no anúncio do Airbnb.
- **numberofreviews**: o número total de comentários que recebeu o anúncio do Airbnb.
- **last_review**: a data da última review que o anúncio do Airbnb recebeu.
- **reviewspermonth**: número médio de avaliações que o anúncio do Airbnb recebe por mês. - **calculatedhostlistings_count**: o número total de listagens que o host tem.
- **availability_365**: o número de dias que o anúncio do Airbnb estará disponível para reserva em um ano



2.1.1 Verificar e alterar o tipo de dados

1. Tabela de Dimensão: `hosts`

Coluna	Tratamento	SQL
<code>host_id</code>	Garantir não nulo. Não deve haver nulos, pois é a chave.	<code>CAST(host_id AS INT64)</code>
<code>host_name</code>	Tratar nulos. Substituir nomes ausentes por 'N/A' e limpar espaços.	<code>TRIM(COALESCE(host_name, 'N/A'))</code>

Passo 1 (Revisado): Limpeza e Tratamento Individual (SQL Corrigido)

1. Tabela de Dimensão: `hosts` (Corrigido)

A chave `host_id` agora usa `SAFE_CAST` e filtramos as linhas onde a conversão falhou (ou seja, onde havia um nome como 'Ociel').

```
-- SQL para a View de HOSTS Limpa
CREATE OR REPLACE VIEW
`my-project-laboratoria.project_rota02.hosts_clean` AS
SELECT
-- 1. Garante que o ID do anfitrião seja um inteiro.
SAFE_CAST(host_id AS INT64) AS host_id,

-- 2. Trata valores nulos no nome, substituindo por 'N/A', e remove espaços em branco extras.
TRIM(COALESCE(nomedohost, 'N/A')) AS host_name
FROM
`my-project-laboratoria.project_rota02.hosts`
-- Remove quaisquer anfitriões sem ID (garantia de integridade)
WHERE SAFE_CAST(host_id AS INT64) IS NOT NULL
GROUP BY 1, 2
```

2. Tabela de Dimensão: `rooms`

Coluna	Tratamento	SQL
<code>id</code>	Chave primária. Garantir não nulo.	<code>CAST(id AS INT64)</code>
<code>name</code>	Tratar nulos e limpar.	<code>TRIM(COALESCE(name, 'No Name'))</code>
<code>latitude / longitude</code>	Garantir tipo float e valores dentro do esperado (para geolocalização).	<code>CAST(latitude AS BIGNUMERIC)</code>
<code>minimum_nights</code>	Garantir positivo. Mínimo de noites deve ser ≥ 1 .	<code>CAST(GREATEST(minimum_nights, 1) AS INT64)</code>

```
-- SQL para a View de ROOMS Limpa
CREATE OR REPLACE VIEW
`my-project-laboratoria.project_rota02.rooms_clean` AS
SELECT
```

```

SAFE_CAST(id AS INT64) AS room_id,
TRIM(COALESCE(name, 'No Name')) AS room_name,
neighbourhood,
neighbourhood_group,

-- Garante o tipo numérico para as coordenadas
CAST(latitude AS BIGNUMERIC) AS latitude,
CAST(longitude AS BIGNUMERIC) AS longitude,

room_type,

-- Garante que o mínimo de noites é pelo menos 1 (Assunção de Negócio)
CAST(GREATEST(minimum_nights, 1) AS INT64) AS minimum_nights
FROM
`my-project-laboratoria.project_rota02.rooms`
-- Filtra quartos sem ID válido ou sem localização
WHERE SAFE_CAST(id AS INT64) IS NOT NULL AND latitude IS NOT NULL AND longitude IS NOT NULL;

```

3. Tabela de Fato: reviews

Coluna	Tratamento	SQL
price	Garantir positivo. O preço deve ser maior que 0.	CAST(price AS BIGNUMERIC)
last_review	Garantir formato data. Converter para DATE .	SAFE_CAST(last_review AS DATE)
reviews_per_month	Tratar nulos. Nulo provavelmente significa 0 reviews/mês.	COALESCE(reviews_per_month, 0.0)
availability_365	Garantir intervalo. Deve estar entre 0 e 365.	GREATEST(0, LEAST(CAST(availability_365 AS INT64), 365))

```

-- SQL para a View de REVIEWS (Fatos) Limpa- VERSÃO FINAL

-- Esta View será a sua Tabela de FATOS no PowerBI
CREATE OR REPLACE VIEW
`my-project-laboratoria.project_rota02.reviews_clean` AS
SELECT
-- Chaves (Relacionamentos)
SAFE_CAST(id AS INT64) AS room_id,
SAFE_CAST(host_id AS INT64) AS host_id,

-- Métricas
CAST(price AS BIGNUMERIC) AS price,
SAFE_CAST(last_review AS DATE) AS last_review,
CAST(number_of_reviews AS INT64) AS number_of_reviews,
COALESCE(reviews_per_month, 0.0) AS reviews_per_month,
CAST(calculated_host_listings_count AS INT64) AS calculated_host_listings_count,

-- Métrica Principal: disponibilidade (limpeza e validação)
GREATEST(0, LEAST(CAST(availability_365 AS INT64), 365)) AS availability_365,

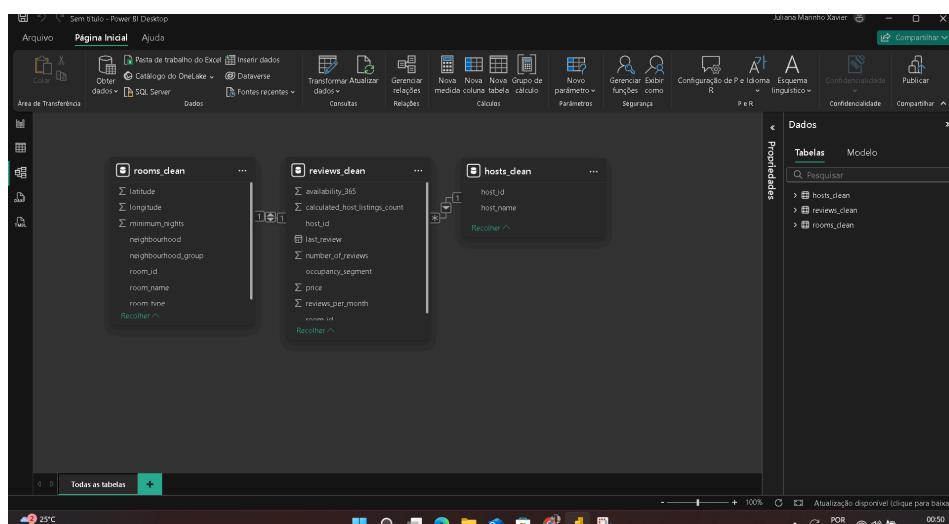
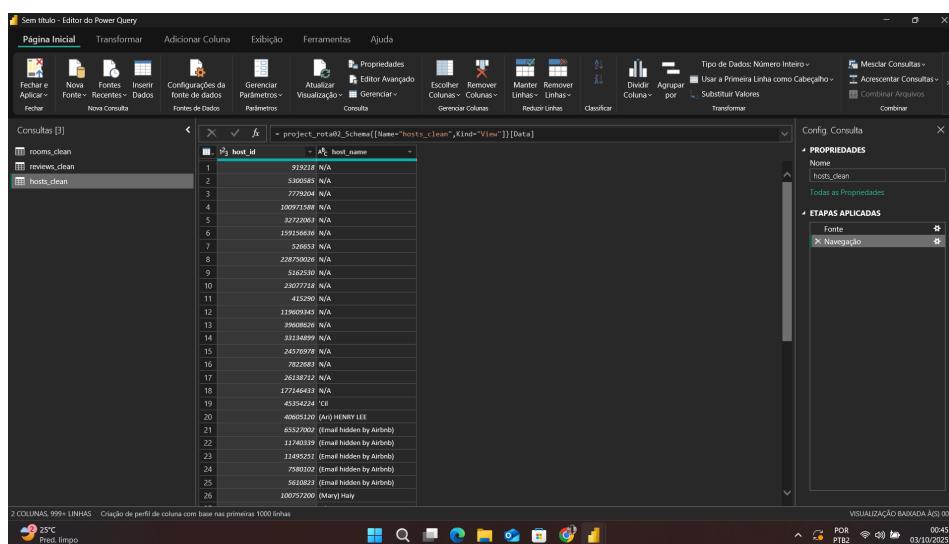
-- Coluna Calculada (Feature Engineering) - REFINADA
CASE
    WHEN GREATEST(0, LEAST(CAST(availability_365 AS INT64), 365)) = 0 THEN '01 - Totalmente Ocupado' -- 10
    0% Ocupado
    WHEN GREATEST(0, LEAST(CAST(availability_365 AS INT64), 365)) <= 90 THEN '02 - Alta Ocupação (Mais de 7
    5% Ocupado)' -- 0 a 90 dias disponíveis
    WHEN GREATEST(0, LEAST(CAST(availability_365 AS INT64), 365)) <= 274 THEN '03 - Ocupação Média' -- 91 a
    274 dias disponíveis
    ELSE '04 - Baixa Ocupação (Ocioso)' -- 275 a 365 dias disponíveis (Mais de 75% Disponível)
END AS occupancy_segment
FROM

```

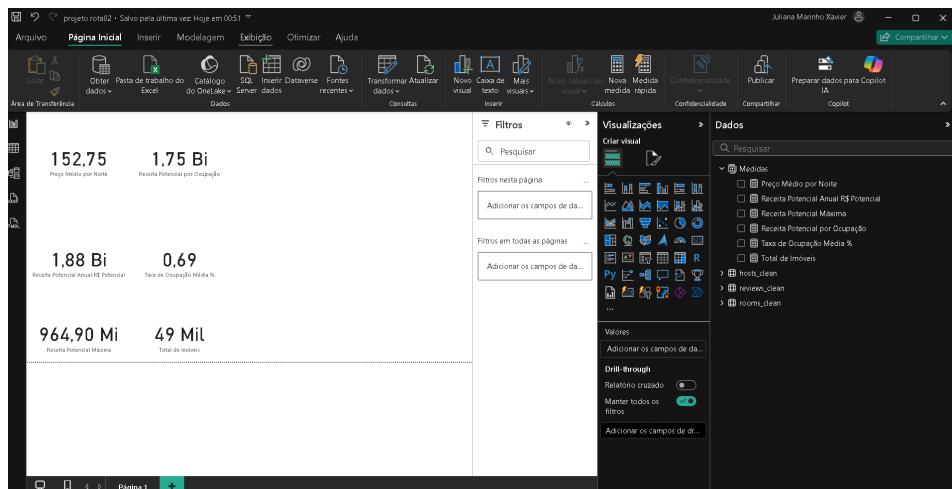
```
'my-project-laboratoria.project_rota02.reviews'  
WHERE  
    SAFE_CAST(id AS INT64) IS NOT NULL AND  
    SAFE_CAST(host_id AS INT64) IS NOT NULL AND  
    price IS NOT NULL AND price > 0;
```



212 Relacionar tabelas



2.1.3 Criar novas variáveis



2.1.4 Utilizar fórmulas DAX

1. Total de Imóveis (KPI Principal)

Para qualquer análise de portfólio, é fundamental saber a contagem total de anúncios.

```
Total de Imóveis = DISTINCTCOUNT('rooms_clean'[room_id])
```

Raciocínio Educativo

- **Por que `DISTINCTCOUNT`?** Embora a Tabela de Fato (`rooms_clean`) tenha uma linha para cada quarto (e o `room_id` seja único nela), usar o `DISTINCTCOUNT` é a melhor prática para contar itens únicos através das relações. Se você filtrar por um bairro, esta função contará corretamente apenas os `room_id`s distintos daquele bairro, utilizando a relação com a dimensão `rooms_clean`.

2. Preço Médio (Métrica de Mercado)

O preço médio é crucial para entender o posicionamento do mercado por região ou tipo de quarto.

```
Preço Médio por Noite = AVERAGE('reviews_clean'[price])
```

Raciocínio Educativo

- **Função `AVERAGE`:** A função `AVERAGE` (média) no DAX é simples, mas poderosa. Ela calcula a média da coluna `price`. Quando você arrastar para um visual no PowerBI e aplicar um filtro (como `neighbourhood_group` da tabela `rooms_clean`), ela automaticamente calculará a média apenas para aquele subconjunto de dados, graças aos **relacionamentos do seu Modelo Estrela**.

3. Taxa Média de Ocupação (%) (KPI de Performance)

Esta métrica mede a eficiência do aluguel, calculando a porcentagem de dias que um quarto esteve potencialmente ocupado (não disponível).

Lembre-se: $\text{Dias Ocupados} = 365 - \text{availability_365}$.

```
Taxa de Ocupação Média % =
VAR Dias_Disponiveis_Total =
    SUM('reviews_clean'[availability_365])
VAR Total_Imoveis_Contados =
    DISTINCTCOUNT('reviews_clean'[room_id])
VAR Dias_Anuais_Media =
    Total_Imoveis_Contados * 365
```

```

VAR Dias_Ocupados_Total =
    Dias_Anuais_Media - Dias_Disponiveis_Total
RETURN
DIVIDE(
    Dias_Ocupados_Total,
    Dias_Anuais_Media
)

```

Raciocínio Educativo

- **Uso de VAR (Variáveis):** Usar variáveis no DAX (com a sintaxe `VAR... RETURN`) é a melhor prática. Aumenta a legibilidade e melhora o desempenho.
- **A Lógica do Cálculo:** Para calcular a média total de ocupação, não podemos simplesmente usar a média de uma coluna. Precisamos somar todos os dias de ocupação e dividir pelo total de dias possíveis.
 1. Calculamos os `Dias_Anuais_Media` (Total de Imóveis * 365 dias).
 2. Calculamos os `Dias_Ocupados_Total` subtraindo os dias disponíveis do total de dias anuais.
 3. A função `DIVIDE` (uma divisão segura) calcula a proporção, nos dando a taxa de ocupação.

4. Coluna Calculada DAX: Potencial Máximo de Hóspedes (Por Quarto)

O objetivo é: **Capacidade Total de Hóspedes** = Capacidade (em dias)/Mínimo de Noites.

Aqui, estamos assumindo que o potencial máximo é 365 dias (todos os dias do ano).

```

Potencial Max Hóspedes Anual =
365 / 'rooms_clean'[minimum_nights]

```

room_id	room_name	neighbourhood	neighbourhood_group	latitude	longitude	room_type	minimum_nights	Potencial Max Hóspedes Anual
56559	City Room - Private & Cozy BedRoom	Harlem	Manhattan	40.71219	-73.94993	Private room	1	365
58059	Furnished Room on Historic Cooper Hill	Harlem	Manhattan	40.82311	-73.94937	Private room	1	365
63666	City Room - Private Luxury Suite	Harlem	Manhattan	40.8118	-73.94484	Private room	1	365
219892	Lovely Bdr in Harlem, Manhattan	Harlem	Manhattan	40.8176	-73.95541	Private room	1	365
224004	Artly Loft-like Harlem Apartment	Harlem	Manhattan	40.80468	-73.95261	Private room	1	365
257461	Spacious room in beautiful apt!	Harlem	Manhattan	40.81672	-73.94857	Private room	1	365
281651	Beautiful Private Bedroom - Downstairs	Harlem	Manhattan	40.80571	-73.95882	Private room	1	365
343971	Beautiful Room in a Beautiful New NYC Apartment	Harlem	Manhattan	40.81777	-73.95	Private room	1	365
449120	Your Own Private Entrance Studio in Stylish Duplex	Harlem	Manhattan	40.8179	-73.9419	Private room	1	365
470582	PRIVATE Room in Spacious, Quiet Apt	Harlem	Manhattan	40.82537	-73.9482	Private room	1	365
578162	Sunny Private Bedroom by Express Train, Collegiate	Harlem	Manhattan	40.81799	-73.93116	Private room	1	365
729496	WEST HARLEM PRIVATE BEDROOM	Harlem	Manhattan	40.80517	-73.93404	Private room	1	365
733264	I Pvt. Room in Upper West Manhattan	Harlem	Manhattan	40.81959	-73.94775	Private room	1	365
795136	Beautiful Bedroom in Harlem	Harlem	Manhattan	40.81724	-73.94183	Private room	1	365
803468	Spacious Private Room in Master Bedrm PRIVATE	Harlem	Manhattan	40.82574	-73.94183	Private room	1	365
831117	Private bedroom in Manhattan	Harlem	Manhattan	40.82324	-73.9401	Private room	1	365
903941	Beautiful Bedroom in Manhattan	Harlem	Manhattan	40.81724	-73.93833	Private room	1	365
933972	Great Bedroom in Manhattan	Harlem	Manhattan	40.82058	-73.9425	Private room	1	365
1151741	Large, room in Z1 on park w/ no discounts	Harlem	Manhattan	40.80449	-73.95907	Private room	1	365
1181326	Comfortable Central Park North Rm	Harlem	Manhattan	40.8004	-73.95645	Private room	1	365
1342315	Furnished Room + Bath in Hamilton Heights	Harlem	Manhattan	40.82243	-73.94507	Private room	1	365
1632420	NYC Harlem cozy private room	Harlem	Manhattan	40.81307	-73.93854	Private room	1	365
1791105	Harlem Oasis	Harlem	Manhattan	40.81757	-73.9335	Private room	1	365
1817829	Enjoy Harlem Renaissance Sunny Room	Harlem	Manhattan	40.82519	-73.9453	Private room	1	365
1979766	Full of Harlem Soul & Charm	Harlem	Manhattan	40.82458	-73.9452	Private room	1	365
2262853	Paddy Pad	Harlem	Manhattan	40.80719	-73.93293	Private room	1	365

5. Medida DAX: % de Quartos Disponíveis por Bairro

Esta é uma **Medida DAX** que compara o número de quartos disponíveis com o total de quartos em um determinado contexto (bairro, tipo de quarto, etc.).

Usaremos a métrica `availability_365` da tabela de Fato (`V_CLEAN_REVIEWS`). Para simplificar, vamos definir um quarto como "Disponível" se tiver `availability_365 > 0`.

```

% Quartos Disponíveis (por Contexto) =
VAR Total_Imoveis =
    DISTINCTCOUNT('reviews_clean'[room_id])
VAR Imoveis_Disponiveis =
    CALCULATE(
        DISTINCTCOUNT('reviews_clean'[room_id]),
        'reviews_clean'[availability_365] > 0
    )

```

```

)
RETURN
DIVIDE(
    Imoveis_Disponiveis,
    Total_Imoveis
)

```

Raciocínio Educativo

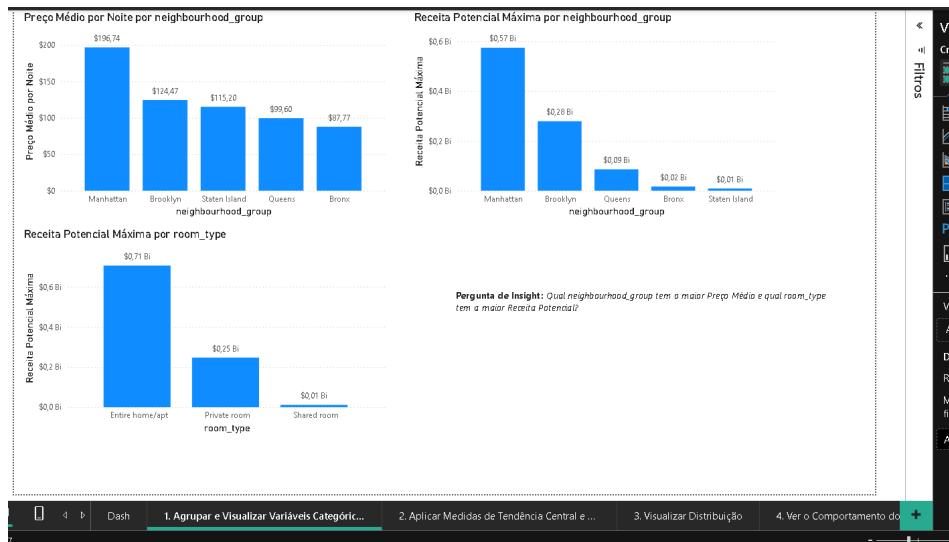
1. **VAR Total_Imoveis** : Usa o `DISTINCTCOUNT` (que criamos anteriormente) para contar todos os quartos visíveis no contexto atual (por exemplo, todos os quartos em "Manhattan").
2. **VAR Imoveis_Disponiveis** : Aqui usamos a função `CALCULATE`, que é a função mais poderosa do DAX. Ela pega o cálculo do `DISTINCTCOUNT` e o executa sob um **novo filtro de contexto**: `'V_CLEAN_REVIEWS'[availability_365] > 0`. Isto significa que ela conta apenas os quartos que têm alguma disponibilidade.
3. **DIVIDE** : Simplesmente divide a contagem dos imóveis disponíveis pelo total de imóveis, resultando na porcentagem.



2.2 Fazer uma análise exploratória

- Agrupar dados de acordo com variáveis categóricas

- Visualizar as variáveis categóricas

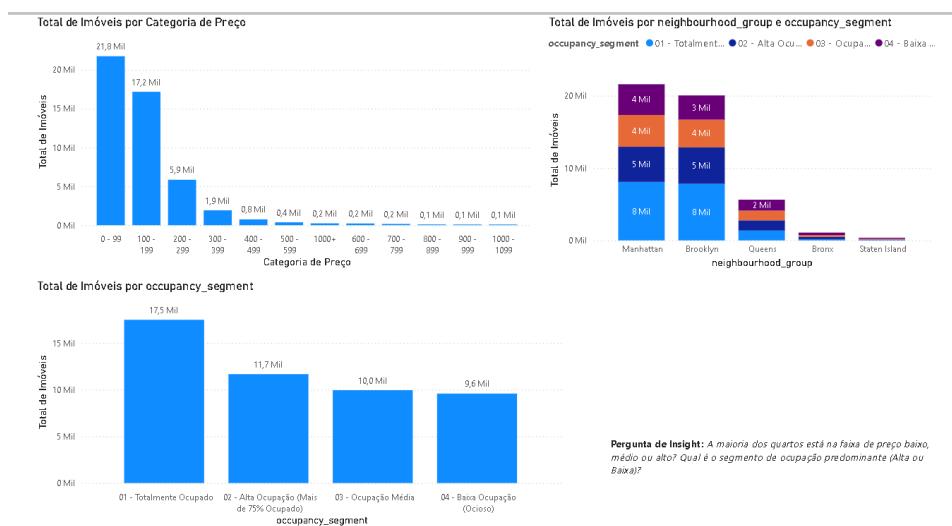


- Aplicar medidas de tendência central

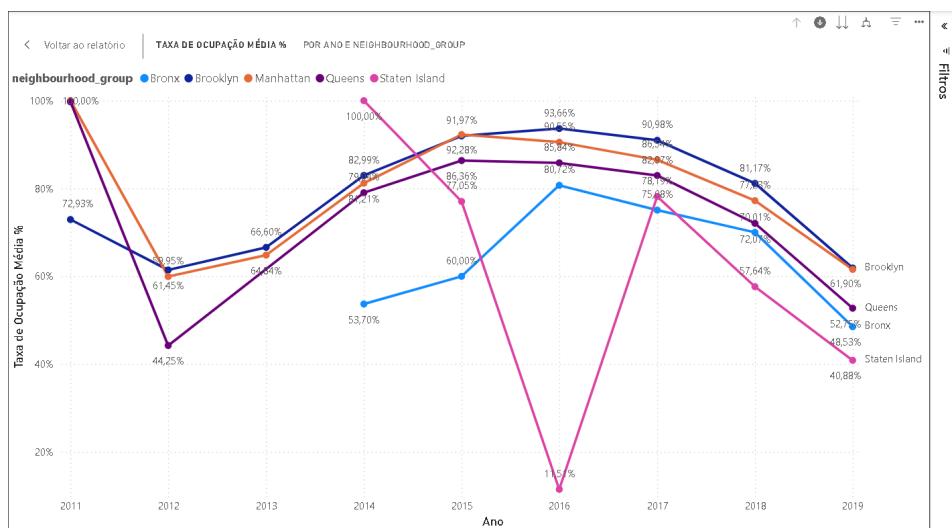
- Aplicar medidas de dispersão

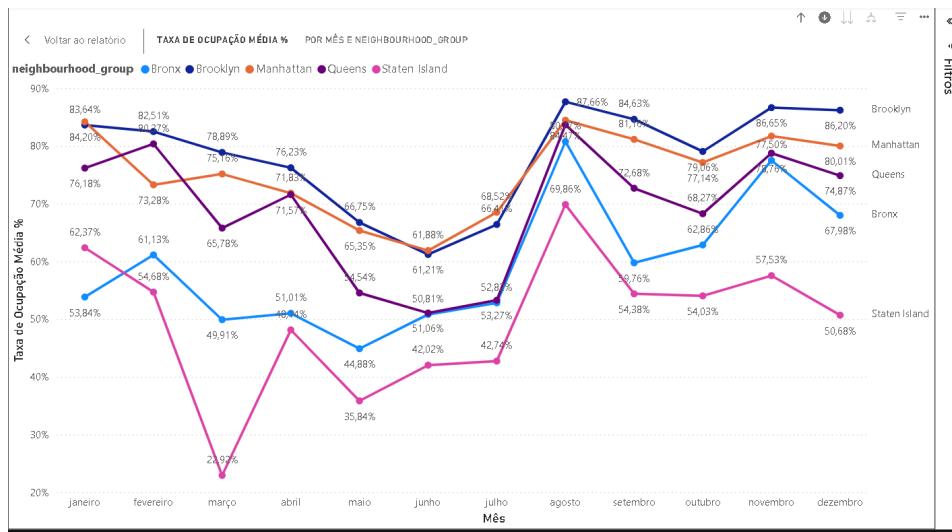


Visualizar distribuição

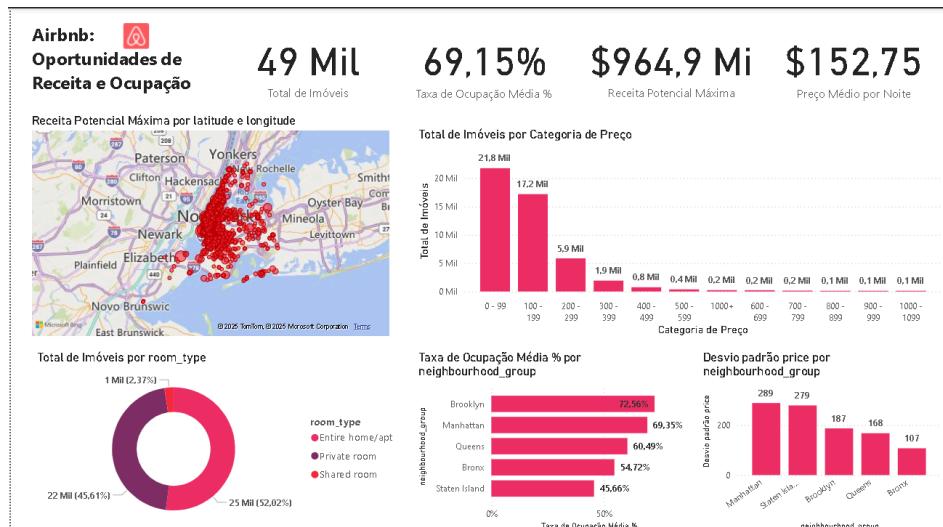


Ver o comportamento dos dados ao longo do tempo





2.3.1 Criar um dashboard com Data Storytelling



[projeto rota02.pbix](#)



2.4 Apresentar resultados

Sumário Executivo do Dashboard Airbnb: Oportunidade e Eficiência

O **Dashboard Airbnb** foi criado com o objetivo de fornecer aos anfitriões e à plataforma uma **visão estratégica** para otimizar a disponibilidade de quartos, maximizar a receita potencial e aprimorar a eficiência operacional. Utilizando princípios de **Business Intelligence (BI)**, este painel transforma dados brutos em *insights*acionáveis.

1. Metodologia e Modelo de Dados

Detalhe	Descrição
Ferramentas	Google BigQuery (SQL) e PowerBI (DAX).

Modelo de Dados	Implementação de um Schema Estrela no PowerBI, onde as tabelas de Dimensão (rooms_clean, hosts_clean) filtram a Tabela de Fato (reviews_clean).
Tratamento de Dados	As três tabelas foram tratadas individualmente no BigQuery (SQL), com foco na conversão segura de tipos de dados (<code>SAFE_CAST</code>) e na filtragem de <i>outliers</i> de preço e disponibilidade.
Métricas Chave (DAX)	Receita Potencial Máxima, Taxa de Ocupação Média (%), Desvio Padrão do Preço e Segmentação de Ocupação (coluna calculada no SQL).

2. Principais Insights e Recomendações Estratégicas

O dashboard permite a análise de tendências geográficas e de preço, levando às seguintes conclusões:

A. Análise de Performance Geográfica

- **Potencial de Receita:** O **Mapa Geográfico** revela que a maior concentração de **Receita Potencial Máxima** está localizada em Manhattan. Os pontos maiores no mapa indicam áreas de alto valor onde o foco na gestão de preços deve ser intensificado.
- **Ocupação vs. Localização:** A **Taxa de Ocupação Média** é mais alta em Brooklyn, sugerindo uma demanda mais consistente, embora com preços potencialmente mais baixos.
- **Volatilidade de Preço (Dispersão):** O gráfico de **Desvio Padrão do Preço** mostra que Manhattan é a área com a maior dispersão de preços. Isso indica um mercado com grande variação entre acomodações de luxo e de baixo custo, sinalizando oportunidades para anfitriões que conseguem se posicionar nos extremos.

B. Análise de Precificação e Inventário

- **Posicionamento de Mercado:** O histograma de **Distribuição de Preço** demonstra que a maioria dos imóveis está concentrada na faixa de \$0 - \$99.
- **Inventário:** A análise do **Tipo de Quarto** mostra que Entire home/apt compõe a maior parte do portfólio.

3. Recomendações para a Tomada de Decisão

1. **Otimização de Preço:** Utilizar o *insight* de **Desvio Padrão do Preço** para ajustar dinamicamente as tarifas, especialmente nas áreas de alta dispersão, explorando a elasticidade de preços.
2. **Foco na Disponibilidade:** Investir em estratégias para aumentar a disponibilidade (`availability_365`) nas áreas de **Alta Ocupação**, convertendo a demanda existente em receita potencial.
3. **Exploração de Nicho:** Investigar as áreas de **Baixa Ocupação/Disponível** (segmento calculado) para entender se a baixa performance se deve à saturação ou a problemas específicos do anúncio.

Este dashboard serve como uma base robusta para análises futuras e para transformar a gestão de aluguel em uma atividade orientada por dados.