

Faculdade de Informática e Administração Paulista - FIAP  
BANCO DE DADOS

**RELATÓRIO: Enterprise Challenge - Gerando  
oportunidades a partir do dados**

Ana Laura - RM89121; Elaine - RM89060; Guilherme - RM88611; Henrique -  
RM88568; Leonardo - RM89107.

São Paulo, Junho de 2022

## RESUMO

Apresentaremos as soluções das necessidades solicitadas da empresa Jk Control. Definiremos e justificaremos nossas escolhas para realização dos desafios. Apresentaremos análises e dashboard como complemento da solução.

### **Alunos:**

Ana Laura F. K. de Macedo  
Elaine dos Santos Silva  
Guilherme da Silva Candido  
Henrique Carvalho Feliciano  
Leonardo Moura

**Área do conhecimento:** Banco de dados (Arquitetura), análise de dados, cloud.

## SUMÁRIO

[RESUMO](#)

[SUMÁRIO](#)

[INTRODUÇÃO](#)

[CONCLUSÃO](#)

## INTRODUÇÃO

As solicitações foram divididas em 3 desafios, no primeiro desafio apresentamos a justificativa da nossa escolha de assets, ou seja, maior volume de entradas e saídas de veículos do condomínio exibida por dia da semana, quinzenalmente, mensalmente, trimestralmente, semestralmente e anualmente. Apresentamos a instalação e configuração do banco de dados NoSQL em nuvem e por fim apresentamos os resultados. Concluimos o primeiro desafio com uma análise descritiva baseando-se sobre os dados cadastrados. No segundo desafio criamos uma narrativa por meio de dados e descrevemos o que foi utilizado para deixar os dados adequados para os usuários finais. Também realizamos análise dos dados e descrevemos o resultado e as técnicas de análises abordadas. Para o último desafio criamos um dashboard e esclarecemos as técnicas de visualização de dados utilizada nesta última parte do desafio.

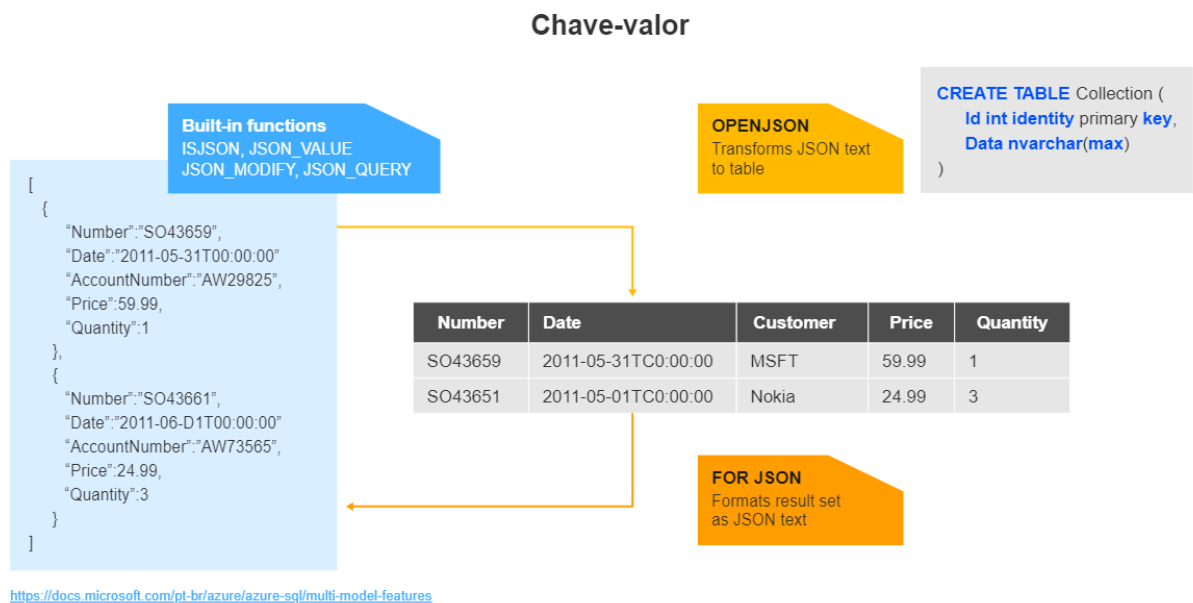
### 1.0 Desafio 1: Surfando na onda de Cloud e NoSQL

**a)** Abordaremos a segunda situação, escolhemos essa opção porque é uma grande solução de entrada para obtermos análises desses dados com alto fluxo em dias comuns. Vale destacar que há períodos sazonais que resultam em uma alta demanda em datas específicas como: datas comemorativas, feriados, eventos do próprio condomínio, etc.

Com esses dados podemos implementar melhorias como maior segurança em determinado horário, investir em um portão automático de melhor performance, aumentar o tamanho da saída ou entrada dependendo do fluxo de carros se mais carros saem ou entram em conjunto em um determinado horário.

**b)** O banco de dados em nuvem escolhido é SQL do Azure, pois processa estruturas não relacionais, tais como: grafos, JSON, pares chave-valor e entre outros.

O modelo de negócio Automação portaria: Projeto Gestão Portaria , armazena dados não estruturados como imagens e vídeos.



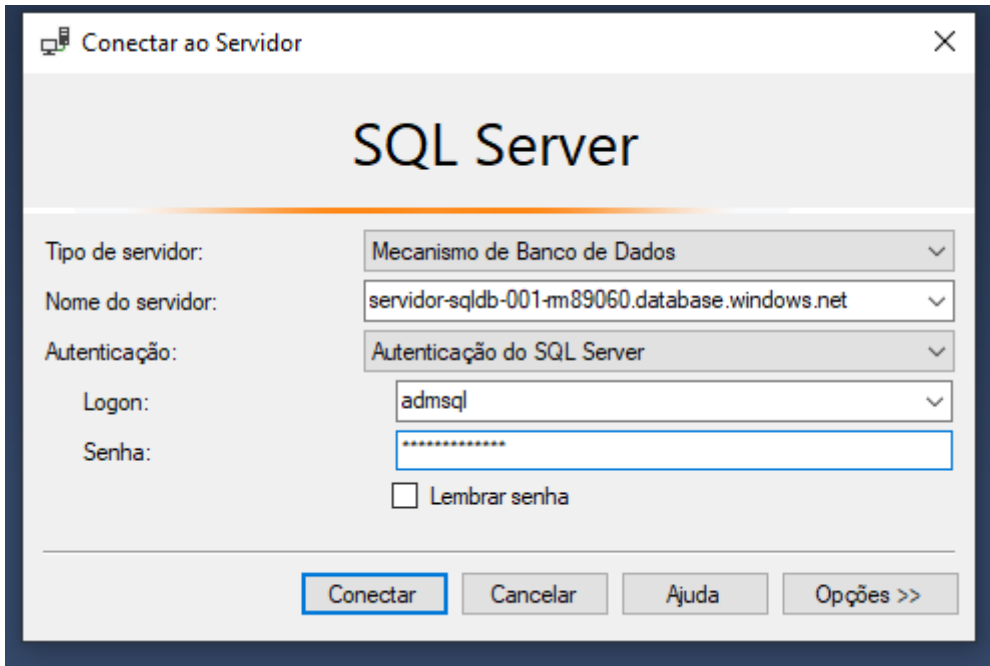
Os Bancos de Dados multi modelos permitem armazenar e trabalhar com dados representados em vários formatos como dados relacionais, grafos, documentos JSON/XML.

- Você tem algumas informações ou estruturas que são mais adequadas para modelos NoSQL e não deseja usar um banco de dados NoSQL separado.
- A maioria dos seus dados é adequada para o modelo relacional e você precisa modelar algumas partes de seus dados no estilo NoSQL.
- Você deseja aproveitar a rica linguagem Transact-SQL para consultar e analisar dados relacionais e NoSQL e integrá-los a uma variedade de ferramentas e aplicativos que podem usar a linguagem SQL.

c) Criação da nova estrutura de dados dentro do NoSQL, utilizando as boas práticas.

- **Conexão com Banco de Dados Azure**

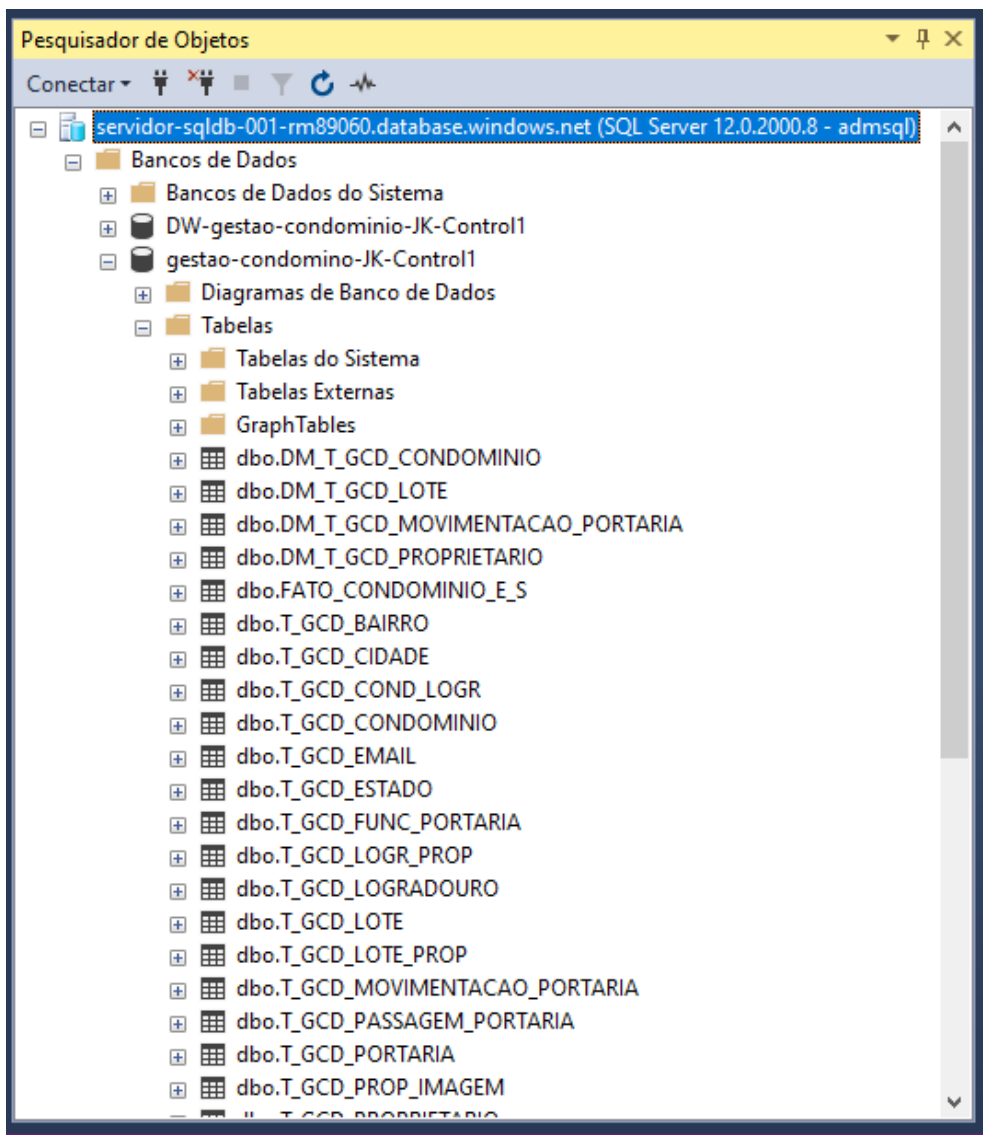
- 1) Após a escolha do Banco de Dados Cloud foi feita a criação do Banco de Dados em nuvem (Azure) e a conexão com o client SSMS.



O Script CREATE Table disponibilizado não estava compatível com o SQL do Azure pois estava com sintaxe Oracle. Foi necessário converter a estrutura do script para adequar com a estrutura da Azure (SQL Server). Depois da criação das tabelas foram feitos os INSERT.

**d) Criação de uma massa com dados significativos a partir dos dados do modelo relacional e realização do upload para dentro do banco de dados NoSQL.**

- **Criação na base de Dados / Upload para dentro do banco de dados NoSQL**

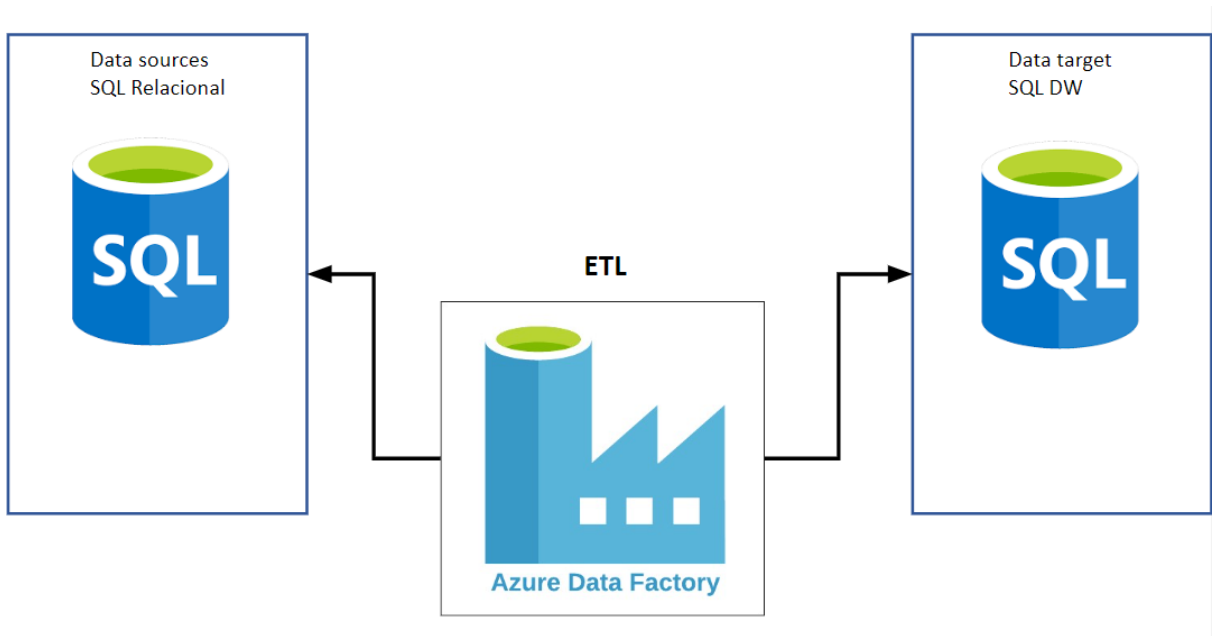


## - Geração Dados Analíticos

Os dados analíticos foram gerados em uma base de dados DW da Azure . O modelo dimensional criado no DW foi schema, onde foi gerado um script CREATE para criar as tabelas FATO e DIMENSÕES.

Abaixo uma visão das ferramentas e processos utilizados.





## - Processo ETL

Abaixo uma visão das etapas utilizadas no processo de ETL para fazer a carga na base de dados do DW.

The screenshot shows the Microsoft Azure Data Factory interface. The pipeline 'pipeline1' is displayed in the 'Validar' (Validate) state. The pipeline consists of five script activities: 'INSERT DM PROPRIETARIO', 'INSERT DM MOVIMENTACAO...', 'INSERT DM LOTE', 'INSERT DM CONDOMINIO', and 'INSERT DM TEMPO'. The pipeline is shown in the 'Validar' (Validate) state. The 'Atividades' (Activities) pane on the left shows the 'Script' activity selected. The 'Saída' (Output) pane at the bottom shows the execution details for the 'INSERT FATO' activity, including the pipeline ID, execution start time, duration, status (Bem-sucedido), and runtime.

| Nome        | Tipo   | Início da execução         | Duração  | Status       | Runtime de   |
|-------------|--------|----------------------------|----------|--------------|--------------|
| INSERT FATO | Script | 2022-06-12T22:12:28.058197 | 00:00:04 | Bem-sucedido | AutoResolvei |

A imagem acima mostra a criação do pipeline com cinco processos, cada processo executa uma query script. Essa query script contém uma função select que extrai os dados do banco de origem/relacional e realiza um insert no banco de destino (DW).

Processos executados com sucesso:

Microsoft Azure | etl-gestao-condominios

pipeline1

Atividades

Validar Depurar Gatilho (1)

Parâmetros Variáveis Configurações Saída

ID de execução de pipeline: e8834b38-f12d-411e-9b3e-f691dd056a8f

| Nome                     | Tipo   | Início da execução         | Duração  | Status       | Runtime de int  |
|--------------------------|--------|----------------------------|----------|--------------|-----------------|
| INSERT FATO              | Script | 2022-06-12T22:12:28.058197 | 00:00:04 | Bem-sucedido | AutoResolveinte |
| INSERT DM TEMPO          | Script | 2022-06-12T22:12:24.186835 | 00:00:04 | Bem-sucedido | AutoResolveinte |
| INSERT DM CONDOMINIO     | Script | 2022-06-12T22:12:20.229855 | 00:00:04 | Bem-sucedido | AutoResolveinte |
| INSERT DM LOTE           | Script | 2022-06-12T22:12:16.177488 | 00:00:04 | Bem-sucedido | AutoResolveinte |
| INSERT DM MOVIMENTACAO P | Script | 2022-06-12T22:12:12.049912 | 00:00:04 | Bem-sucedido | AutoResolveinte |
| INSERT DM PROPRIETARIO   | Script | 2022-06-12T22:12:08.458521 | 00:00:03 | Bem-sucedido | AutoResolveinte |

Acionamento de gatilho onde carrega os dados no DW:

Microsoft Azure | etl-gestao-condominios

pipeline1

Atividades

Validar Depurar Gatilho (1)

Parâmetros Variáveis Configurações Saída

ID de execução de pipeline: e8834b38-f12d-411e-9b3e-f691dd056a8f

| Nome                     | Tipo   | Início                     |
|--------------------------|--------|----------------------------|
| INSERT FATO              | Script | 2022-06-12T22:12:28.058197 |
| INSERT DM TEMPO          | Script | 2022-06-12T22:12:24.186835 |
| INSERT DM CONDOMINIO     | Script | 2022-06-12T22:12:20.229855 |
| INSERT DM LOTE           | Script | 2022-06-12T22:12:16.177488 |
| INSERT DM MOVIMENTACAO P | Script | 2022-06-12T22:12:12.049912 |
| INSERT DM PROPRIETARIO   | Script | 2022-06-12T22:12:08.458521 |

Editar gatilho

Nome \*

CARREGAR DADOS NO DW

Descrição

Tipo \*

ScheduleTrigger

Data de início \*

6/12/22 22:15:00

Fuso horário \*

Tempo Universal Coordenado (UTC)

Recorrência \*

Cada 1 Dia(s)

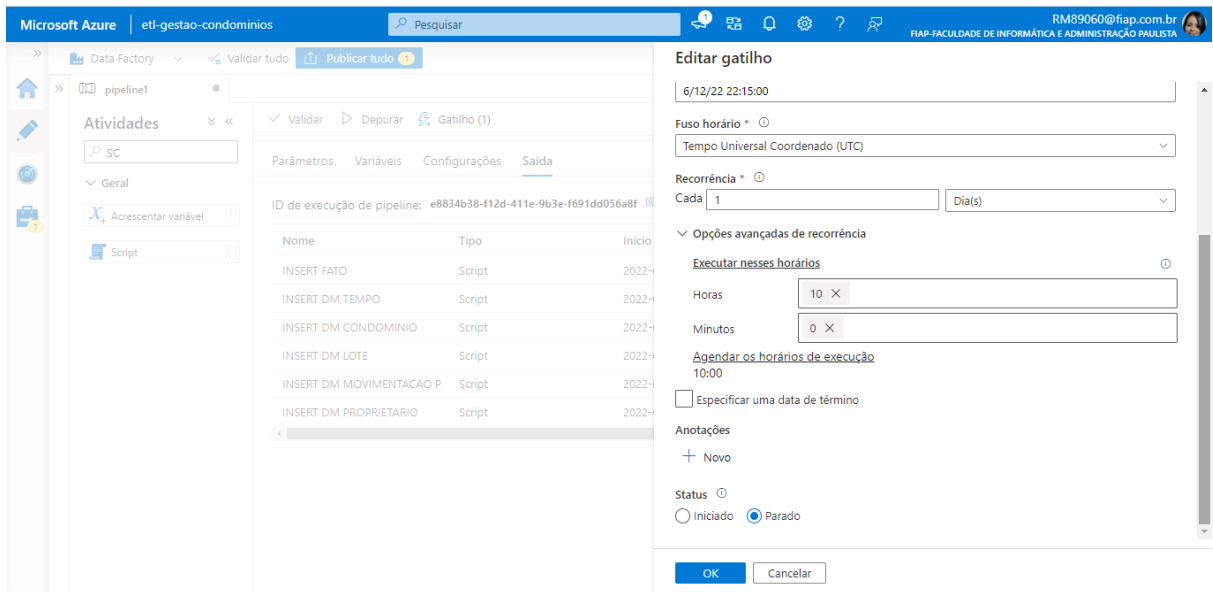
Opções avançadas de recorrência

Executar nesses horários

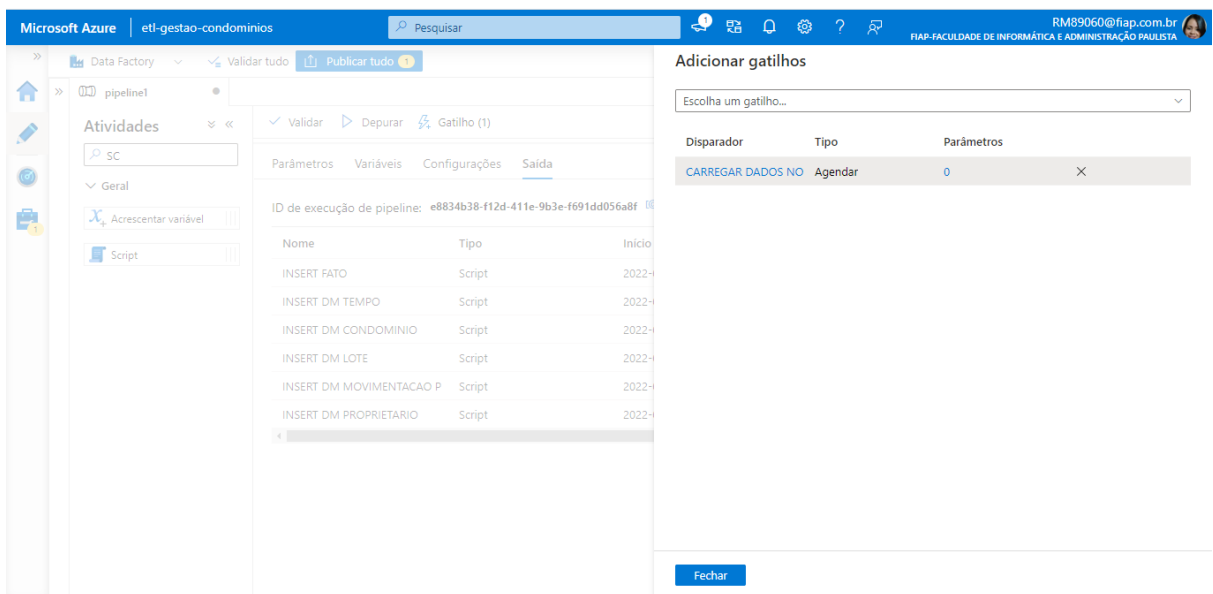
Horas 10

OK Cancelar

Frequência com que o gatilho é disparado:

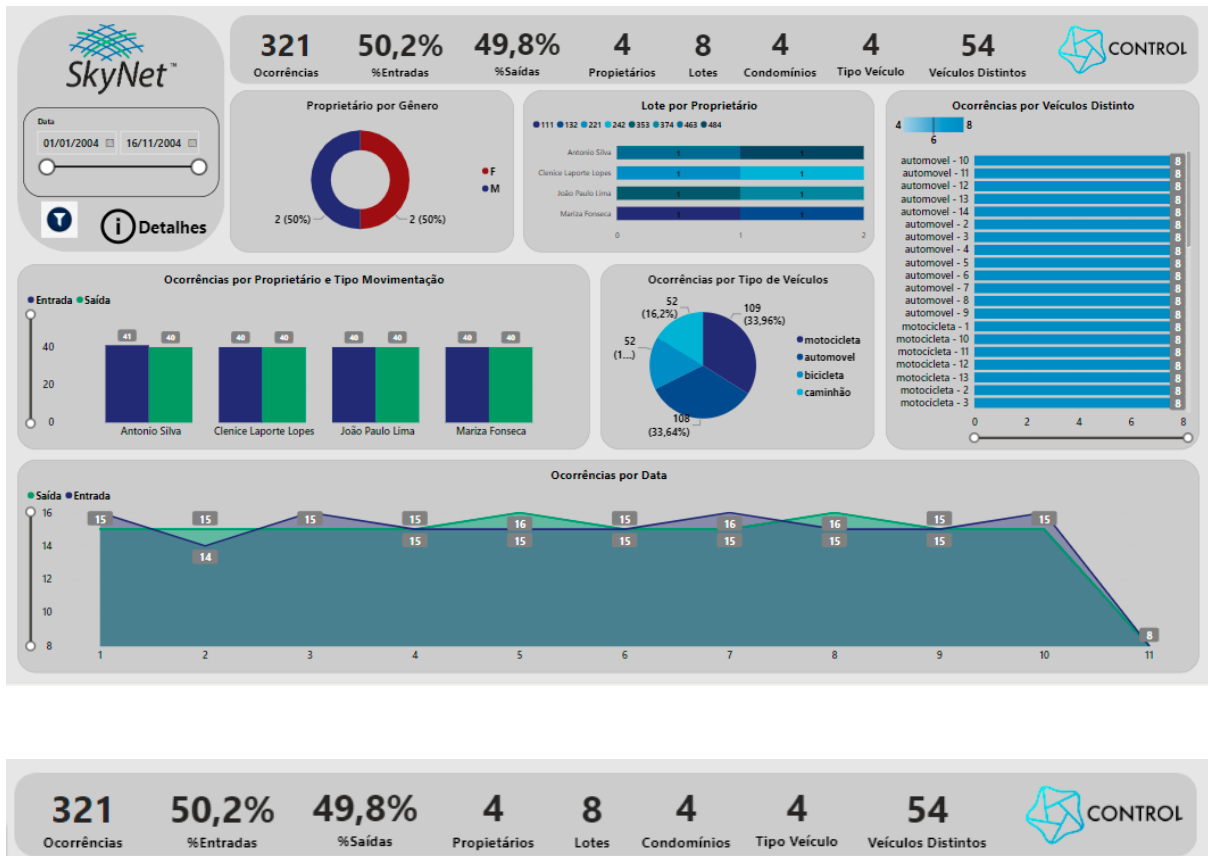


Gatilho ativado com sucesso:

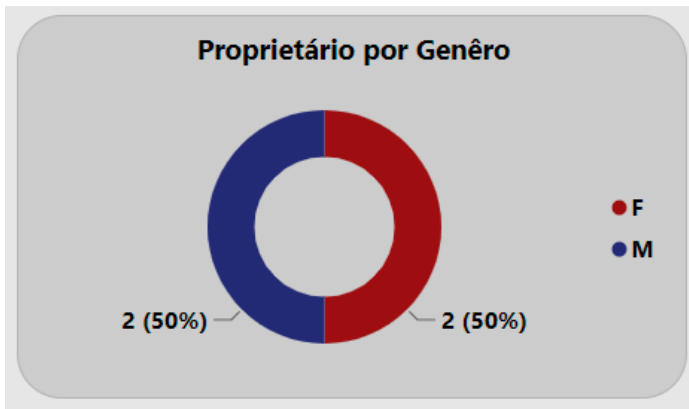


**Observação:** Adicionamos o script gerado em um arquivo .txt chamado '1\_desafio\_scripts\_dbajuniors'.

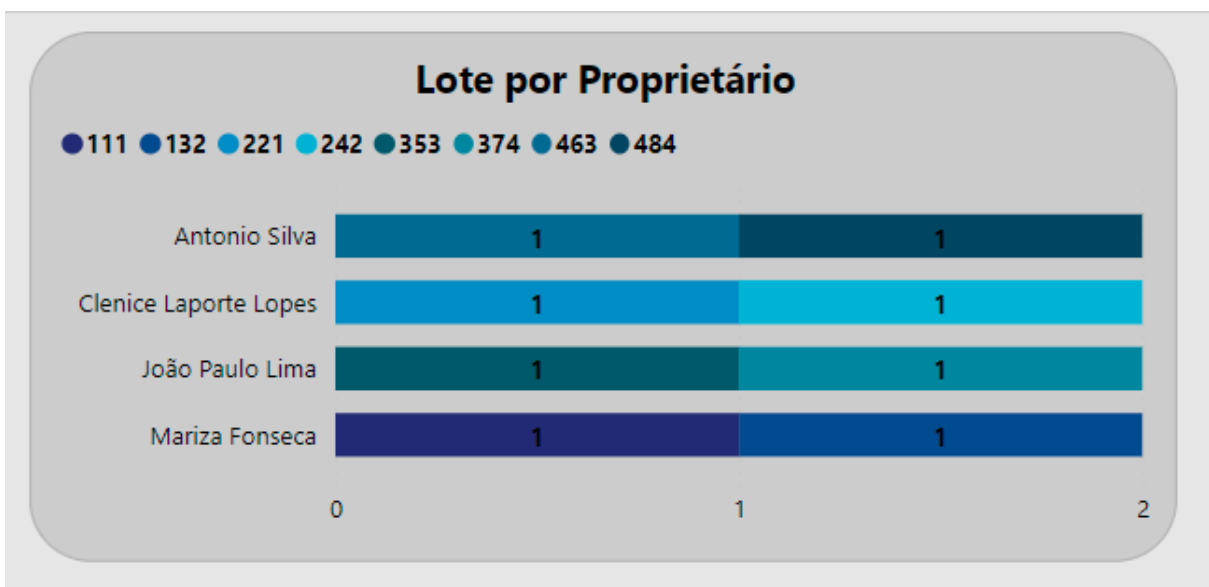
- **Análise descritiva baseando-se nos dados cadastrados**



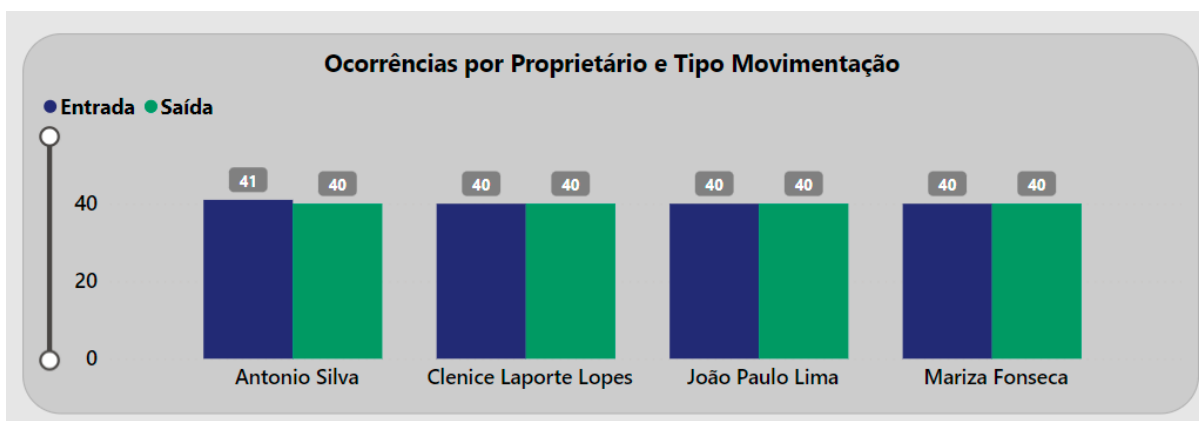
De acordo com a base de dados apresentada em diversos tipos de análises, a JK Control fixou na aba superior o total dos dados abordados como um todo. Neste campo podemos identificar 321 Ocorrências de entrada e saída, com 53 Veículos Distintos, sendo que 50,2% dessas ocorrências são identificadas como Entradas e 49,8% são Saídas. Dadas as ocorrências, também podemos verificar que nessa base de dados, é apresentada 4 (quatro) proprietários, 8 (oito) lotes e 4 (quatro) Condomínios, na qual cada proprietário contém um tipo de veículo e 2 lotes para um único condomínio.



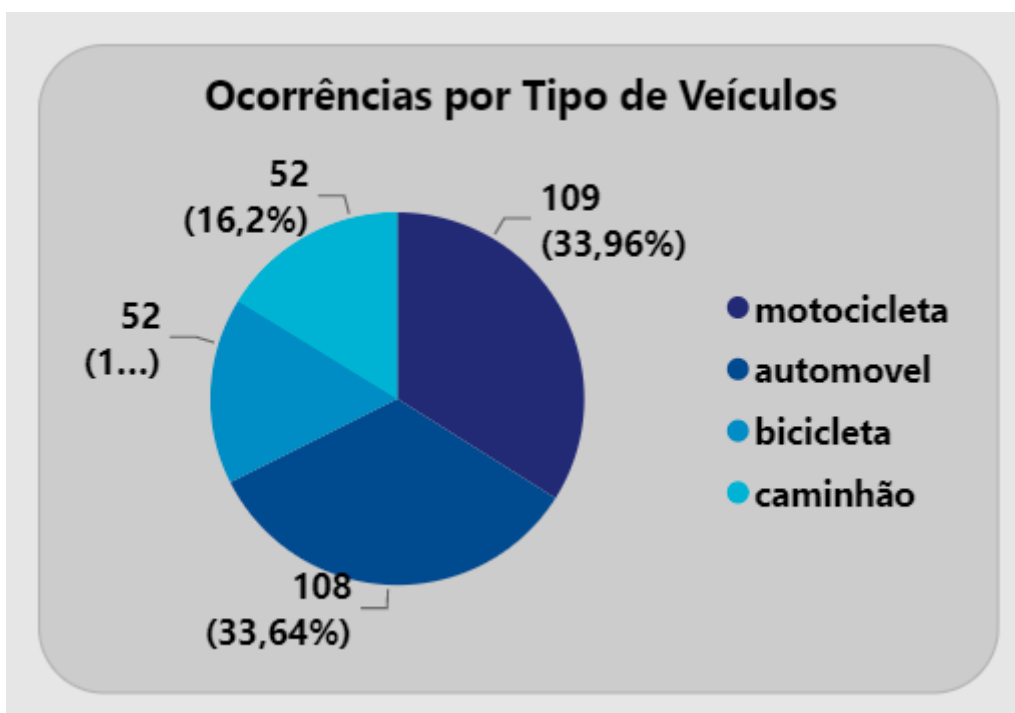
Com os dados gerais abordados no gráfico anterior, podemos visualizar com exclusividade a quantidade de proprietários por gênero, na qual há uma divisão proporcional de 50% dos proprietários serem do gênero Masculino, enquanto os outros 50% eram do gênero feminino.



Dadas as ocorrências do campo superior, podemos verificar que na base de dados, é apresentada 4 proprietários (Antonio Silva, Clenice Laporte Lopes, João Paulo Lima, Mariza Fonseca), 8 lotes (111, 132, 221, 242, 353, 374, 436, 484) e 4 (quatro) Condomínios, na qual cada proprietário contém 2 lotes para um único condomínio.



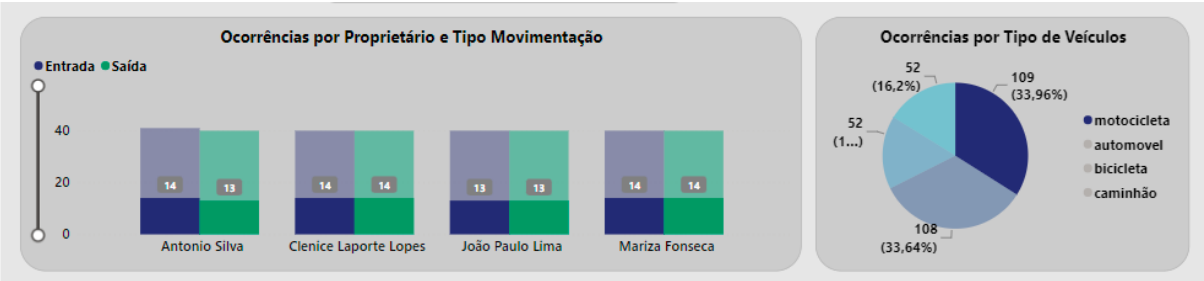
Neste gráfico é destacado o fluxo de ocorrência de cada proprietário. Dentre as 321 ocorrências, 81 ocorrências são do proprietário Antonio Silva (41 entradas e 40 saídas), 80 ocorrências Clenice Laporte Lopes (40 Entradas e 40 Saídas), 80 ocorrências João Paulo Lima (40 Entradas e 40 Saídas), 80 ocorrências Maria Fonseca (40 Entradas e 40 Saídas).



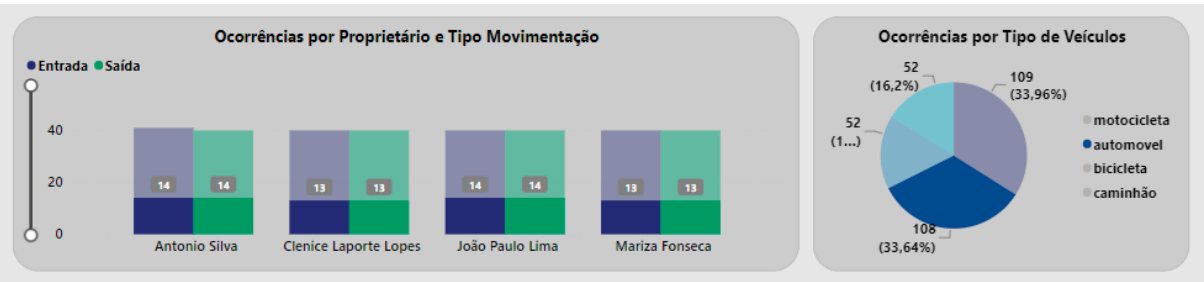
Complementando o Gráfico de Ocorrências por Proprietário e Tipo de Movimentação, temos o gráfico de Tipos de Veículos, na qual podemos filtrar pela ocorrência de cada proprietário, o tipo de veículo que foi registrado na portaria.

Os veículos listados no gráfico, correspondem a um número de ocorrências totais, de cada proprietário, para facilitar a geração de relatórios, por exemplo:

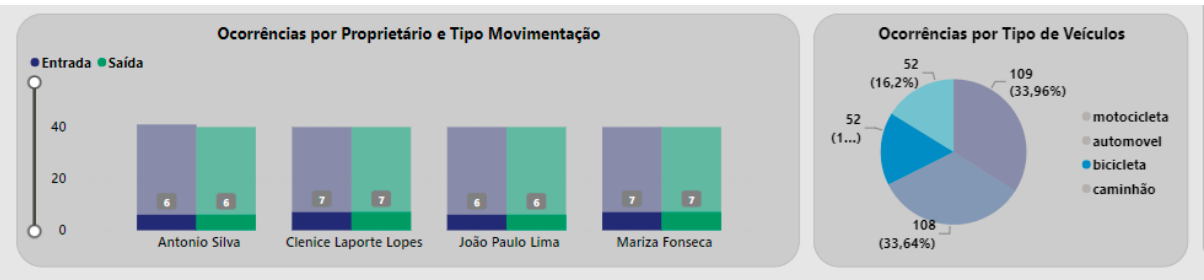
1) Motocicleta → 109 ocorrências (33,96% da soma total de ocorrências).



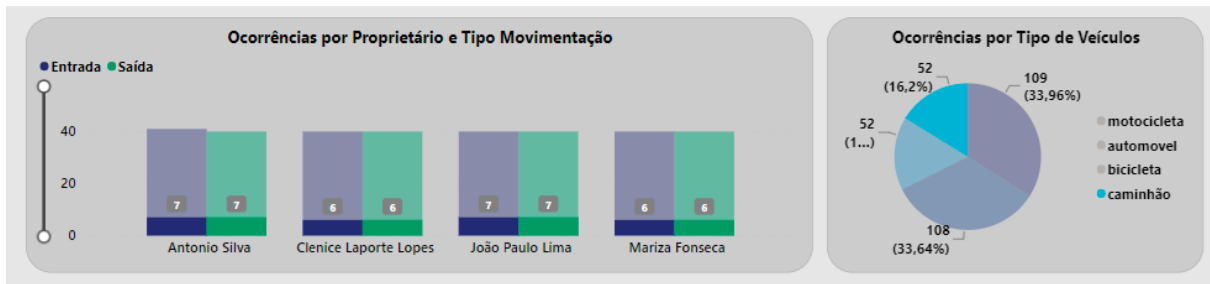
2) Automóvel → 108 ocorrências (33,64% da soma total de ocorrências).



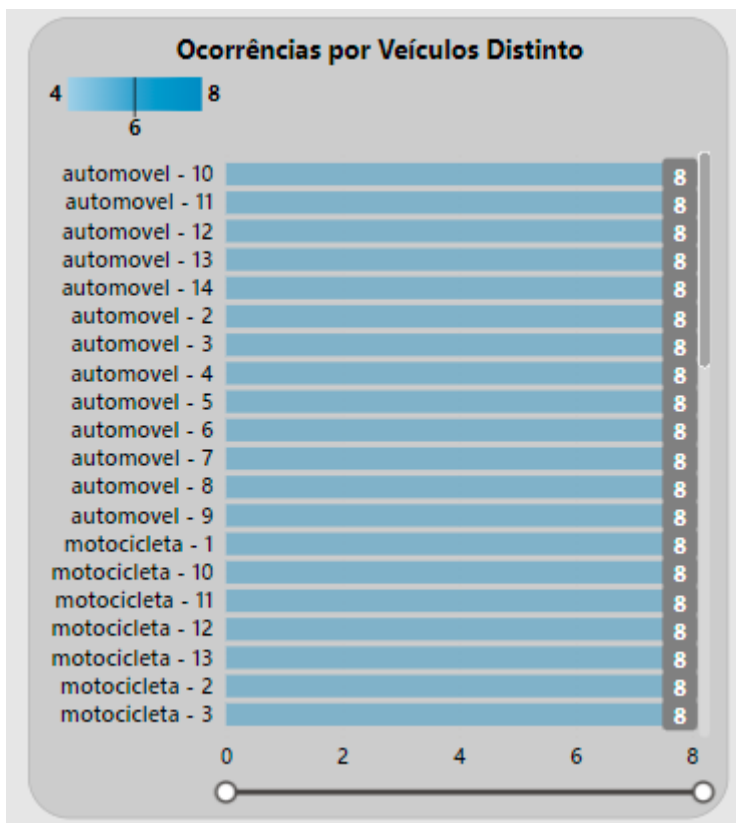
3) Bicicleta → 52 ocorrências (16,2% da soma total de ocorrências)



4) Caminhão → 52 ocorrências (16,2% da soma total de ocorrências)



81 ocorrências são do proprietário Antonio Silva (41 entradas e 40 saídas);  
 80 ocorrências Clenice Laporte Lopes (40 Entradas e 40 Saídas);  
 80 ocorrências João Paulo Lima (40 Entradas e 40 Saídas);  
 80 ocorrências Maria Fonseca (40 Entradas e 40 Saídas).

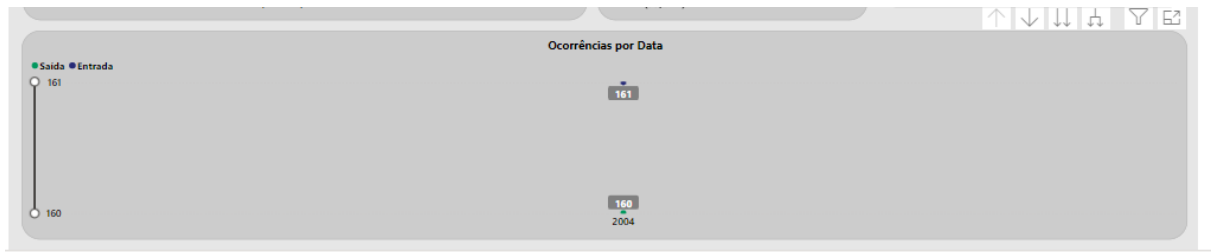


O gráfico destacado fornece a quantidade de veículos com as suas respectivas ocorrências para cada proprietário.

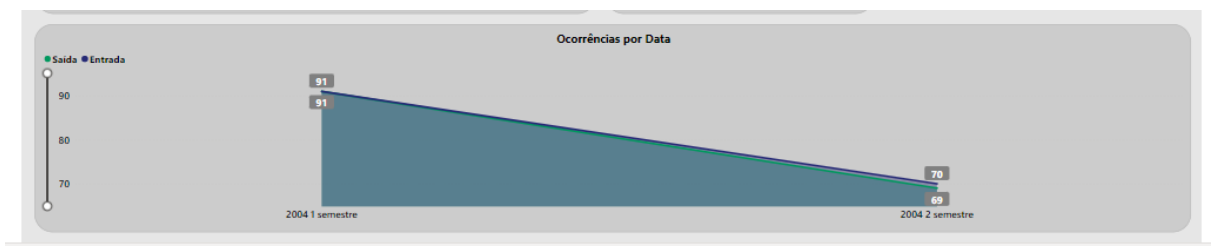
Na parte inferior do dash podemos realizar uma análise por períodos de tempo, realizando um drill down e drill up com os botões de ações, nesta



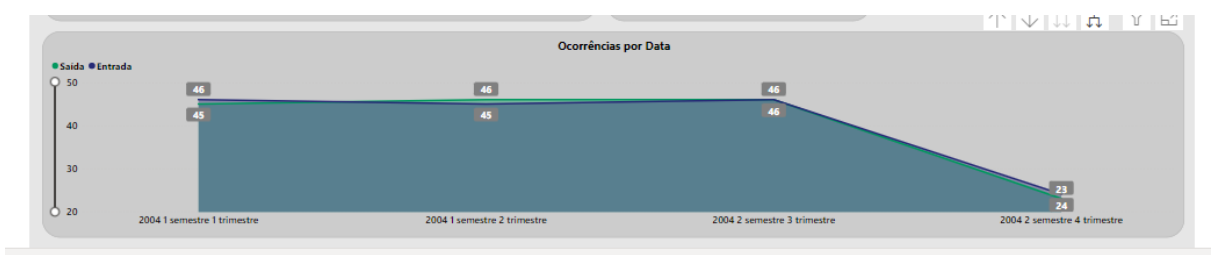
visualização temos o volume de entrada e saída anual, onde foram registradas 161 entradas e 160 saídas no ano de 2004.



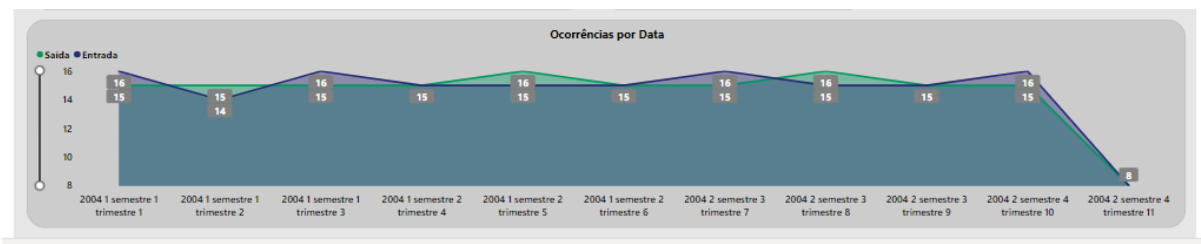
Em seguida realizando um drill down temos uma visão por semestre, percebemos que no primeiro semestre de 2004 houveram mais ocorrências de entrada e saída comparado ao segundo.



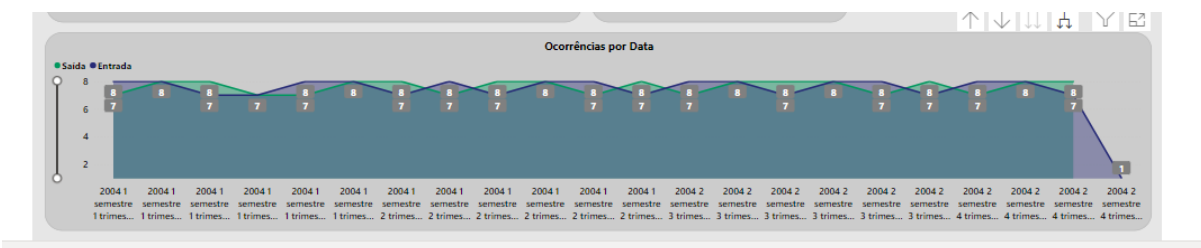
Analisando os dados por trimestre notamos que a queda ocorreu no último trimestre do ano, onde até então nos três primeiros havia uma média de 46 entradas e saídas, podemos analisar um fato interessante no segundo semestre, onde existe uma saída registrada a mais do que as entradas, o contrário do que aconteceu no primeiro semestre, o que pode indicar que um veículo entrou no primeiro semestre e saiu no segundo ou houve falha na segurança, onde perdeu se uma entrada e uma saída de um veículo.



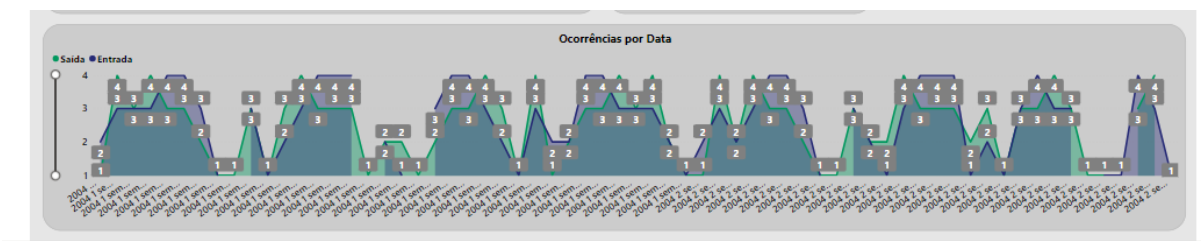
Ao analisar mensalmente observamos que alguns moradores têm o hábito de passar longos períodos fora, visto que em alguns meses existem mais entradas do que saídas e outros mais saídas do que entradas, o que indica que alguém passa um período de um mês fora e volta no mês seguinte.



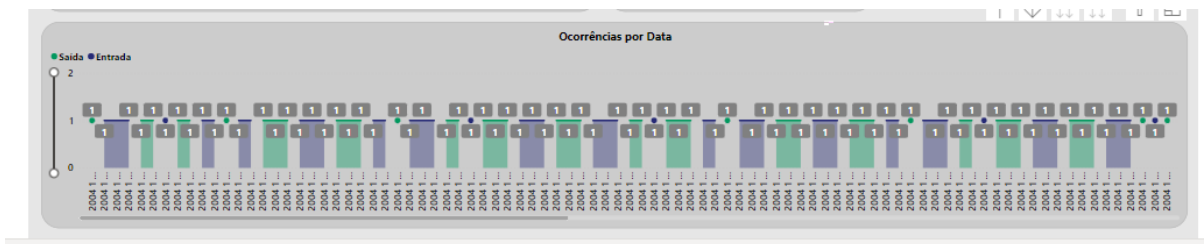
Na análise quinzenal confirmamos a teoria da análise mensal, ontem percebemos que quinzenalmente existem moradores que entram em uma quinzena e saem na outra e vice versa.



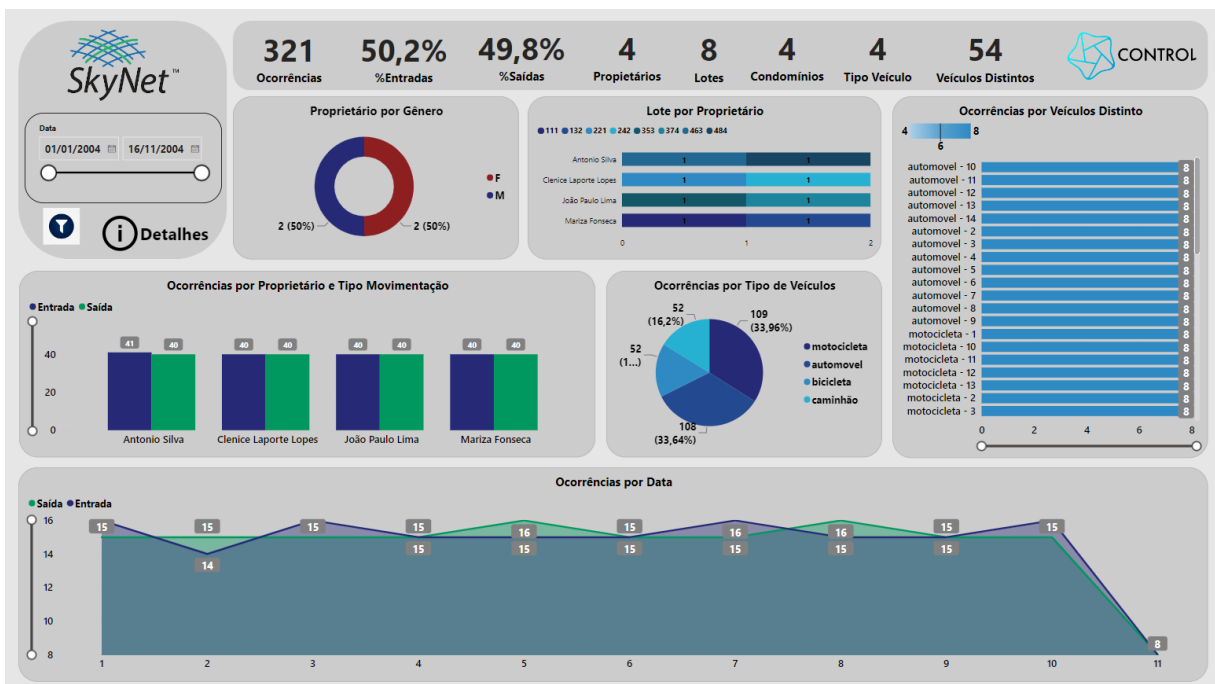
Olhando semanalmente os dados indicam que em algumas semanas não há registros de entrada ou saída, ou somente o registro mínimo.

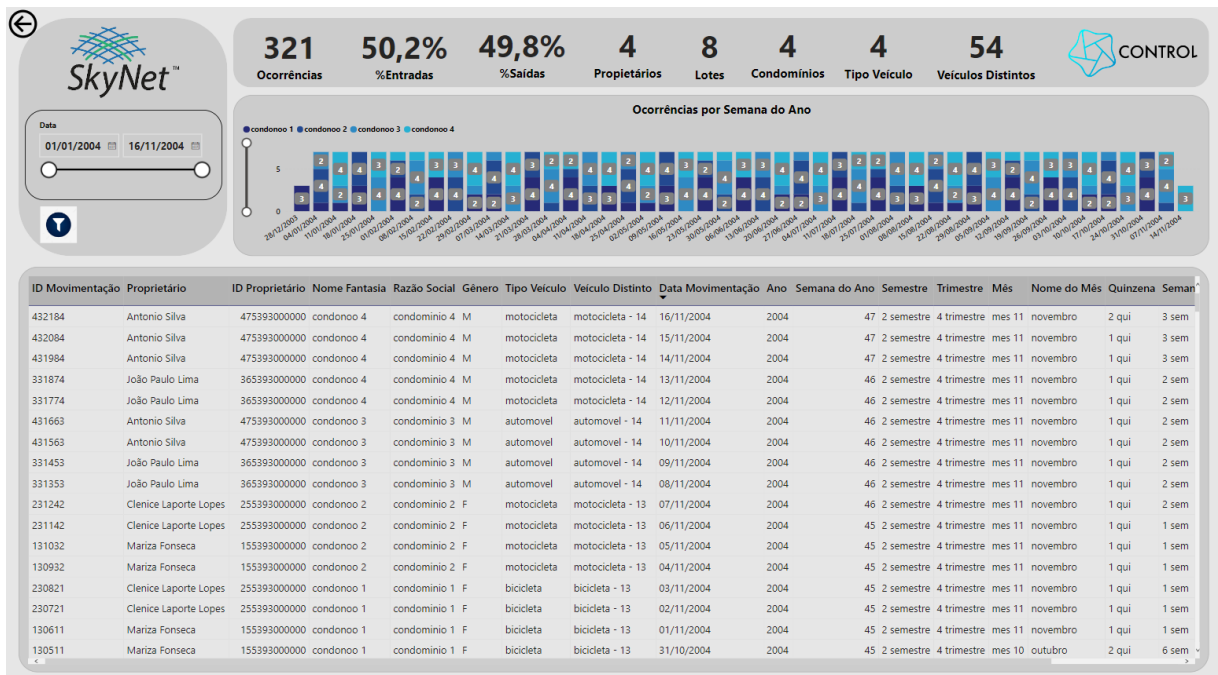
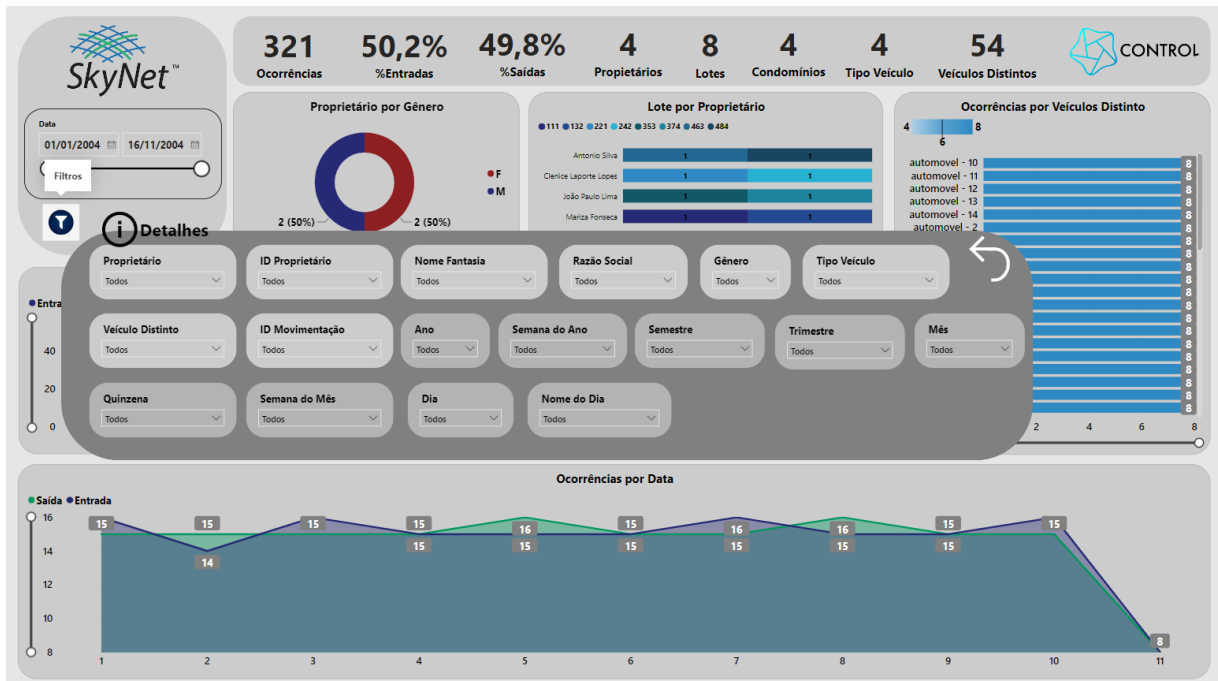


Diariamente temos o registro mínimo de uma entrada e uma saída em média.



## - Dashboard





link para acesso do dashboard:

<https://app.powerbi.com/view?r=eyJrljoiNmNjOTE4NTctN2Q0OS00ZjVmLTg3OGQtMTk2MTM2NmM3YTZkIiwidCI6IjRmYjFkMjM2LTNhNzctNGE4ZS1hN2E4LWExNmVjOGFhNDIjMCJ9>

## 1.2 Desafio 2: Gerando oportunidades com o uso dos dados

A empresa Skynet atua há mais de dez anos na área de desenvolvimento de sistemas, com foco em aplicações web e mobile, obtendo um crescimento exponencial, chegando a atingir mais de mil clientes no país nesses últimos cinco anos.

Um dos proprietários da Skynet, o sr. Freddy Miercoles (apelido Fred), está atuando em um audacioso projeto de automação visando atender milhares de condomínios.

Após uma indicação encontrou a JK Control a parceira ideal para seus projetos envolvendo IOT sendo que, nessa fase do projeto, o foco será a gestão da portaria, que tem como principal objetivo controlar a entrada e saída de pessoas e veículos dentro do condomínio.

A Base de Dados foi criada de acordo com o modelo de negócio Automação portaria: Projeto Gestão Portaria e inserida no Banco de Dados. Depois de alimentar o Banco de Dados foi criado um modelo analítico (DW) a fim de atender a necessidade do negócio.

- **Persistência dos dados:** Os dados são persistidos em uma base de dados sql cloud no Azure.

- **Dados analíticos:** Os dados para realizar a análise foram maior volume de entradas e saídas de veículos do condomínio. Essa informação pode ser exibida por dia da semana, quinzenalmente, mensalmente, trimestralmente, semestralmente e anualmente.

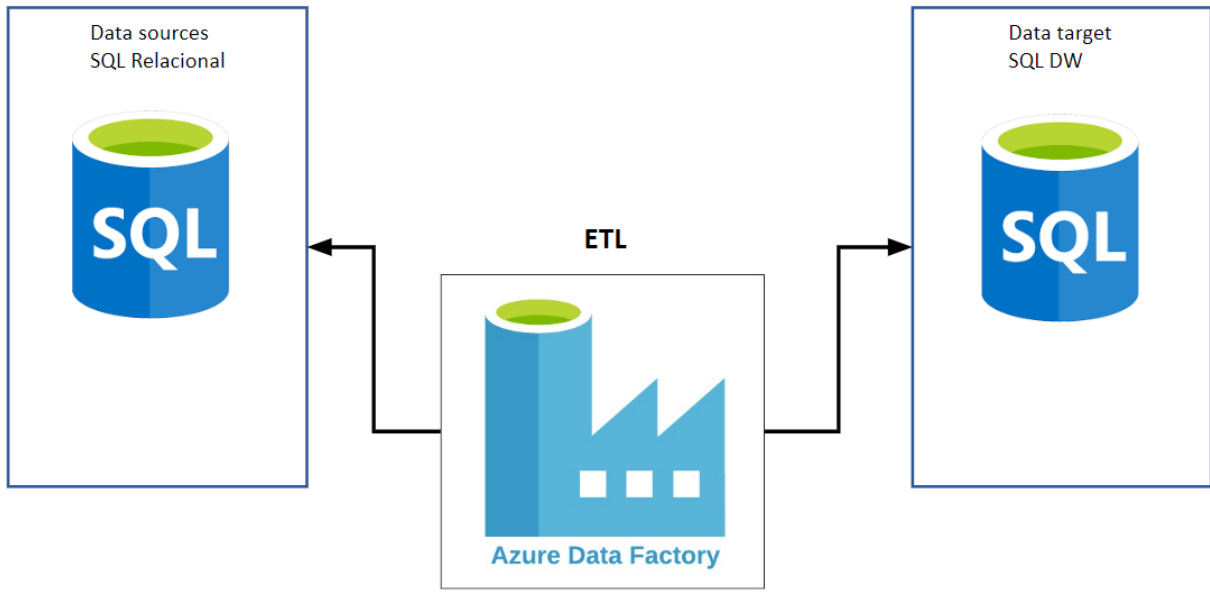
- Para atender esses dados analíticos foi criado um base de dados DW no banco de dados SQL na cloud Azure, com modelo dimensional star schema. O modelo contém as seguintes tabelas:

**Tabelas**                      **dimensão:**                      DM\_T\_GCD\_PROPRIETARIO,  
DM\_T\_GCD\_MOVIMENTACAO\_PORTARIA,                      DM\_T\_GCD\_LOTE,  
DM\_T\_GCD\_CONDOMINIO e TEMPO\_CONDOMINIO\_E\_S.

**Tabela fato:** FATO\_CONDOMINIO\_E\_S.

- **Persistência DW:** Para realizar o processo de ETL na base DW utilizamos o fluxo abaixo. No fluxo temos a tabela de origem no banco de dados SQL na cloud Azure. O data factory é o responsável em extrair os dados da base de dados de origem e persistir na base de dados de destino que é o DW no

banco de dados SQL na cloud Azure.



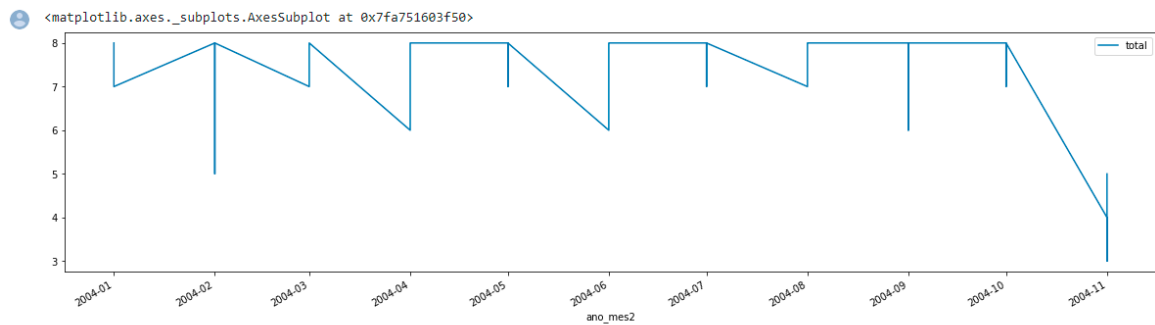
- **Aplicando modelos analíticos:** Através dessa base de dados DW, os dados são exportados para um arquivo .csv e utilizando a ferramenta PANDAS do Python, é aplicado diversos modelos de análise: análise descritiva e preditiva, modelos estacionários.

link colab:

[https://colab.research.google.com/drive/19byeYYcg6mGDdljWb\\_bs0xCHSzNh0IO1?usp=sharing](https://colab.research.google.com/drive/19byeYYcg6mGDdljWb_bs0xCHSzNh0IO1?usp=sharing)

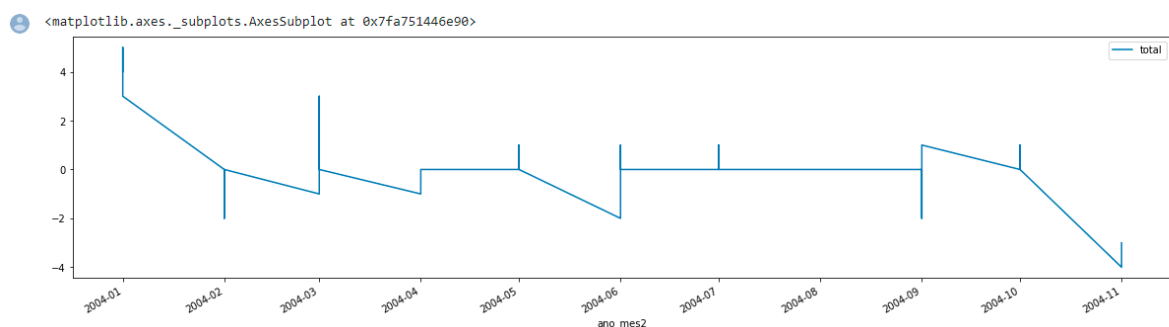
## - Análises

**GRÁFICO DE SÉRIES TEMPORAIS** (Os gráficos de série temporal são criados traçando um valor agregado (uma contagem ou estatística, como soma ou média) em uma linha do tempo. Os valores são agregados usando intervalos de tempo com base no intervalo de tempo nos dados sendo plotados.)

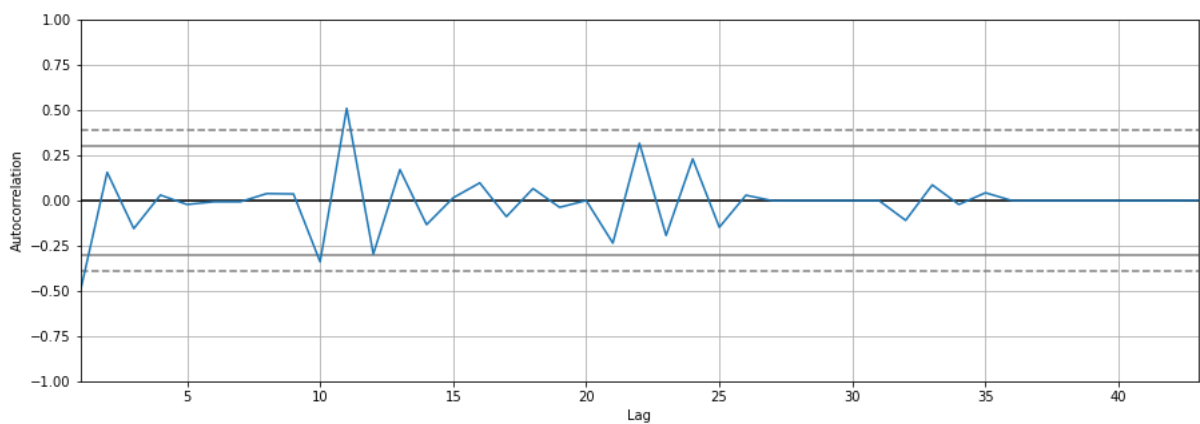


Observando o intervalo de tempo de 11 meses destacado no gráfico, podemos apontar que o número de ocorrências, ao longo da série, possui uma constância relevante de fluxo (entrada e saída) até o mês 10. A partir do mês 11 há uma queda brusca de ocorrências.

**MODELO ARIMA** (um modelo auto-regressivo integrado de médias móveis é uma generalização de um modelo auto-regressivo de médias móveis.)

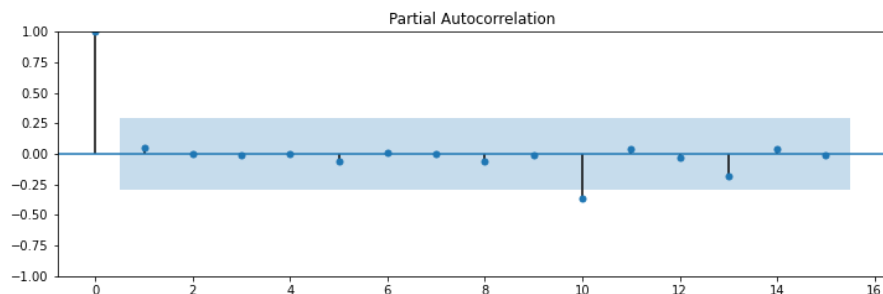
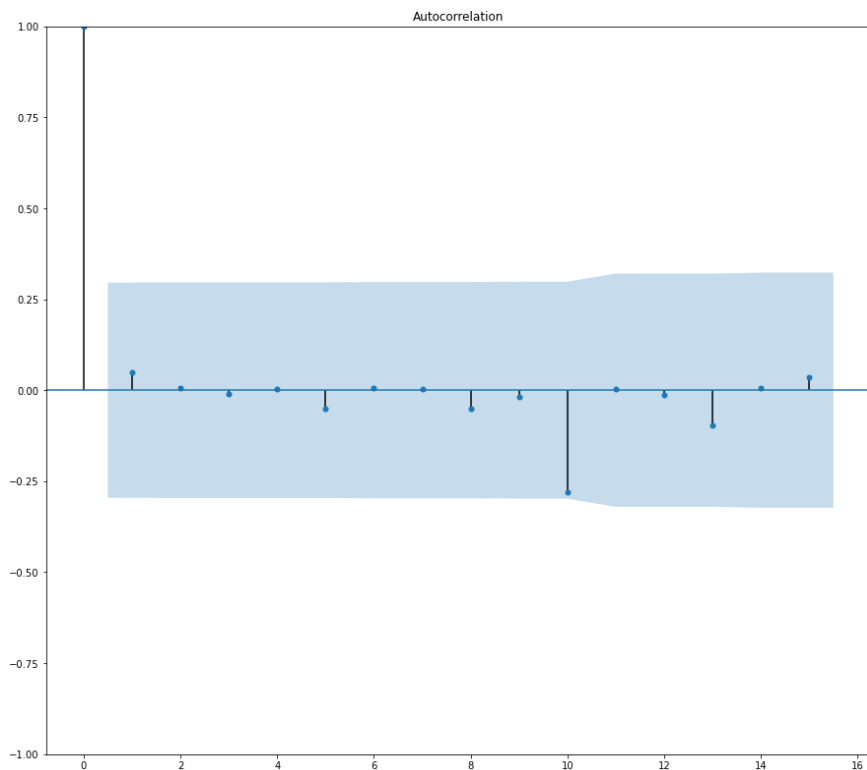


AR(p) - analisando a autocorrelação parcial



Nesse modelo de autocorrelação podemos notar que os dados de série temporal não tem nenhuma anormalidade.

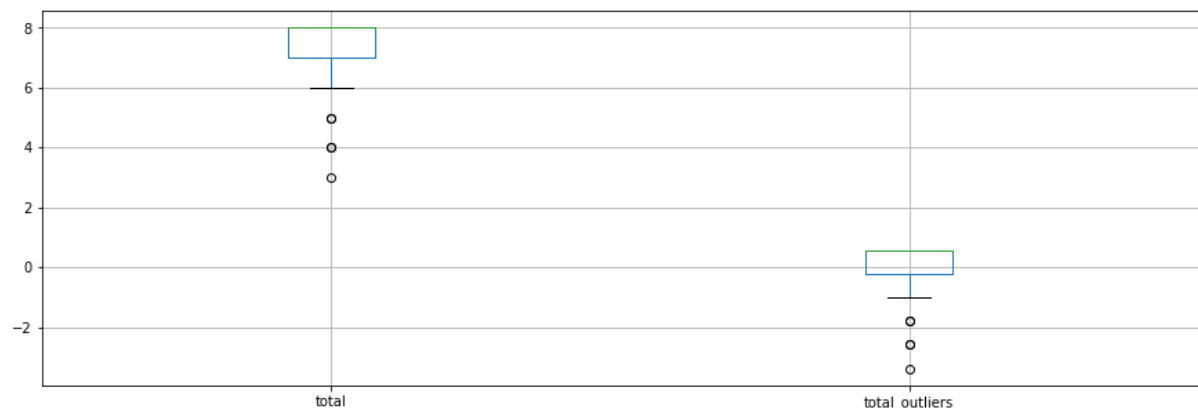
**AUTOCORRELAÇÃO PARCIAL** (A autocorrelação parcial é a correlação entre as observações em uma série temporal que não é contabilizada por todos os intervalos mais curtos entre essas observações.)



Podemos notar na autocorrelação parcial que não há uma variedade maior que -0,25 no eixo Y diretamente no lag 10. Já na autocorrelação segue da mesma forma, não há uma variedade maior que -0,25 no lag 10.



## DADOS OUTLIERS



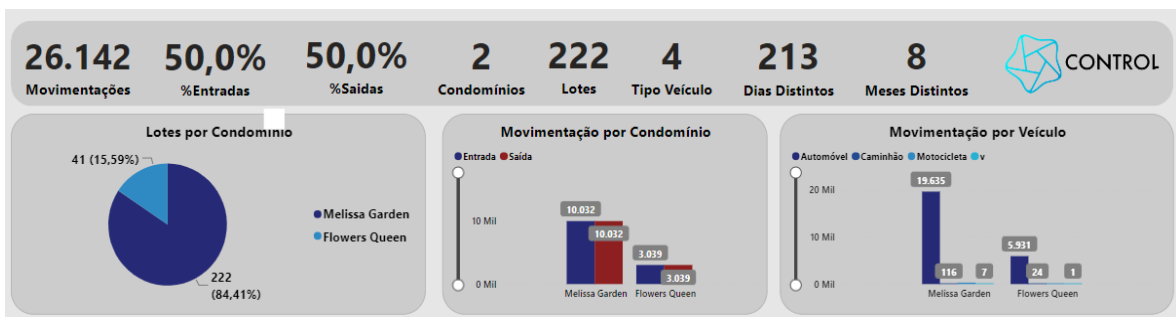
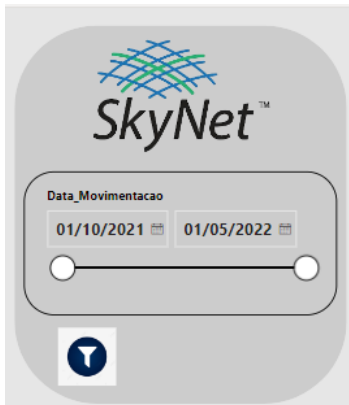
Nessa visão podemos notar que temos pouca divergência da normalidade dos dados da JK Control.

### 1.3 Desafio 3: A magia da visualização de dados

- Esclarecimento técnico das técnicas de visualização:

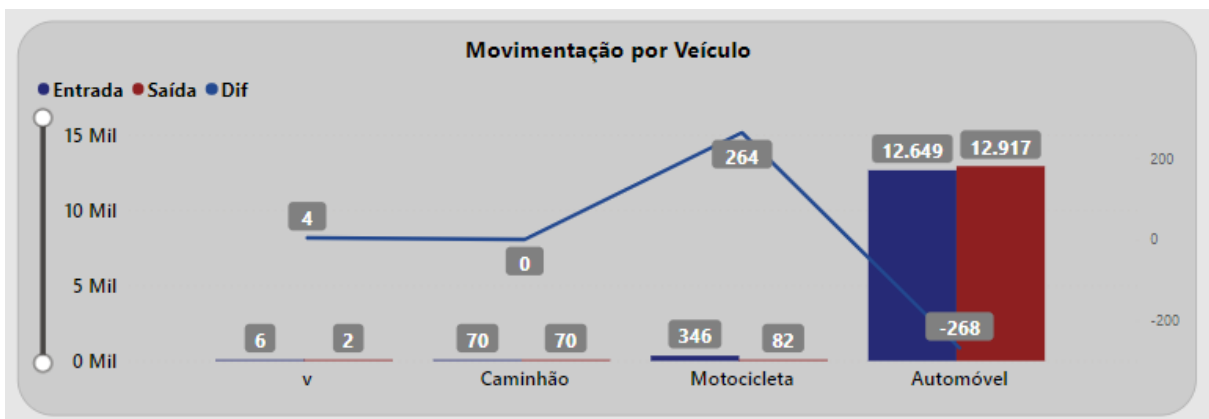
O painel foi desenvolvido na ferramenta Power B.I, utilizando linguagem M no tratamento dos dados e função DAX na criação de cálculos matemáticos. Em relação a visualização respeitamos o espaçamento entre os gráficos, criamos uma padronização de cores, letras, números e logo das empresas envolvidas no projeto.

Ao olhar para uma página, o cérebro humano inicia-se pelo lado superior esquerdo até a parte inferior direita, então os indicadores e filtros encontram-se nessa ordem de grandeza porque são as informações relevantes para entender as movimentações dentro do condomínio. Os filtros estão agrupados em um único ícone, assim ganhamos mais espaço dentro do dashboard e maior dinamismo dos usuários.



Logo abaixo dos indicadores temos um gráfico de pizza mostrando a distribuição de lotes por condomínio, onde a maior concentração está no Melissa Garden. Em seguida, com um gráfico de barras podemos concluir que todos moradores que saíram dos condomínios retornaram. E por último, outro gráfico de barras comparando a movimentação por veículo, onde percentualmente temos a mesma concentração em automóveis.

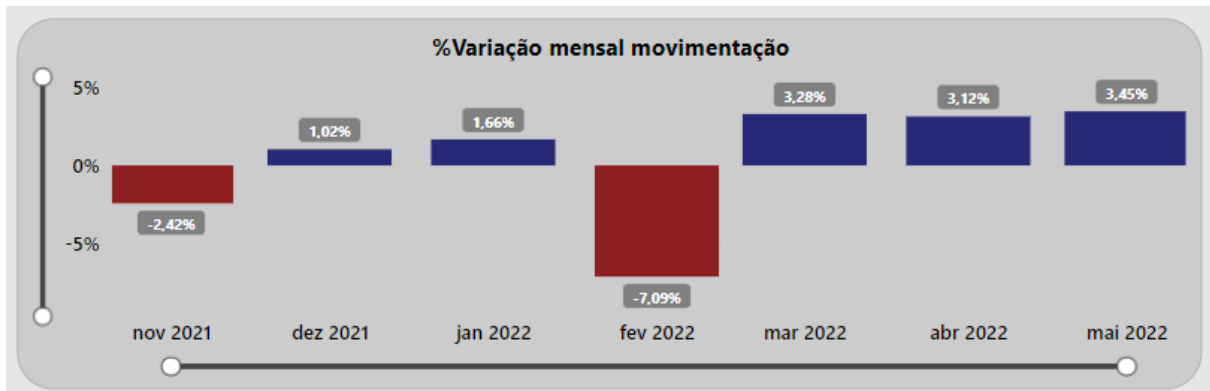
Analisando esses dados gostaríamos de entender se teve algum veículo que saiu mas não retornou ao condomínio ou vice-versa, para isso utilizamos um gráfico de combinação:



Aqui podemos concluir que o tipo de veículo "V" são outliers em relação a

essa base. Também que tivemos mais entradas de moto do que saída (+264) e mais saídas de automóveis do que entrada, tendo um gap de 268 de movimentação.

Gostaríamos de entender a variação mensal da movimentação dentro do condomínio, então comparamos o mês atual vs o mês anterior com um gráfico de barras com eixos negativos e positivos:

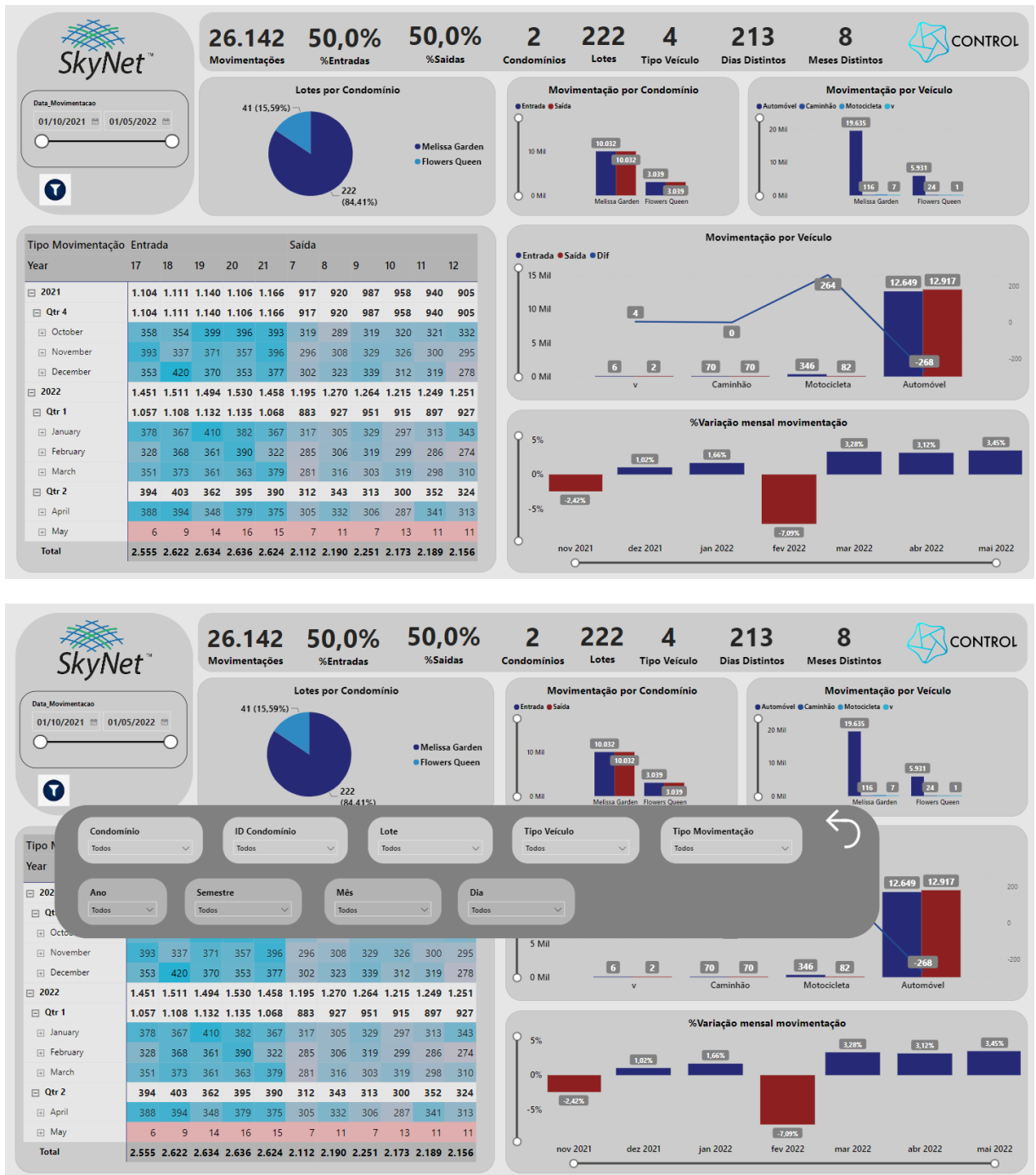


Apenas nov/ 2021 e fev/ 2022 tivemos uma menor movimentação nos condomínios, feriados e férias dos moradores explicaria essa baixa movimentação. Já em maio/ 2022 temos apenas um dia de informação, logo estamos comparando parcialmente com o mês de abril, sendo 120 e 116 movimentações respectivamente. Por último conseguimos distinguir a entrada e saída dos moradores por hora e data, aplicando um drill down e drill up pelas colunas e/ ou linhas para uma melhor análise e entendimento da informação:

| Fazer drill on Linhas |         |       |       |       |       |       |       |       |       |       |       |
|-----------------------|---------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|
| Tipo Movimentação     | Entrada |       |       |       |       | Saída |       |       |       |       |       |
| Year                  | 17      | 18    | 19    | 20    | 21    | 7     | 8     | 9     | 10    | 11    | 12    |
| 2021                  | 1.104   | 1.111 | 1.140 | 1.106 | 1.166 | 917   | 920   | 987   | 958   | 940   | 905   |
| Qtr 4                 | 1.104   | 1.111 | 1.140 | 1.106 | 1.166 | 917   | 920   | 987   | 958   | 940   | 905   |
| October               | 358     | 354   | 399   | 396   | 393   | 319   | 289   | 319   | 320   | 321   | 332   |
| November              | 393     | 337   | 371   | 357   | 396   | 296   | 308   | 329   | 326   | 300   | 295   |
| December              | 353     | 420   | 370   | 353   | 377   | 302   | 323   | 339   | 312   | 319   | 278   |
| 2022                  | 1.451   | 1.511 | 1.494 | 1.530 | 1.458 | 1.195 | 1.270 | 1.264 | 1.215 | 1.249 | 1.251 |
| Qtr 1                 | 1.057   | 1.108 | 1.132 | 1.135 | 1.068 | 883   | 927   | 951   | 915   | 897   | 927   |
| January               | 378     | 367   | 410   | 382   | 367   | 317   | 305   | 329   | 297   | 313   | 343   |
| February              | 328     | 368   | 361   | 390   | 322   | 285   | 306   | 319   | 299   | 286   | 274   |
| March                 | 351     | 373   | 361   | 363   | 379   | 281   | 316   | 303   | 319   | 298   | 310   |
| Qtr 2                 | 394     | 403   | 362   | 395   | 390   | 312   | 343   | 313   | 300   | 352   | 324   |
| April                 | 388     | 394   | 348   | 379   | 375   | 305   | 332   | 306   | 287   | 341   | 313   |
| May                   | 6       | 9     | 14    | 16    | 15    | 7     | 11    | 7     | 13    | 11    | 11    |
| Total                 | 2.555   | 2.622 | 2.634 | 2.636 | 2.624 | 2.112 | 2.190 | 2.251 | 2.173 | 2.189 | 2.156 |

Utilizamos um gradiente de cores para entender a discrepância entre as movimentações, e confirmando a quarta análise, temos mais entradas do que saídas dentro dos condomínios.

#### - Dashboard



link para acesso do dashboard:

<https://app.powerbi.com/view?r=eyJrJoiOWQ3MzNkMjltMjkwYi00NGNiLTg2MGYtOTExNjc2ZjMyYzcyliwidCI6IjRmYjFkMjM2LTNhNzctNGE4ZS1hN2E4LWExNmVjOGFhNDIjMCJ9>

## **CONCLUSÃO**

Por meio da análise dos dados do condomínio, conseguimos obter diversos insights tanto em forma de dashboard utilizando a ferramenta PowerBI quanto analisando dados no python através do Colab, esses poderão auxiliar os administradores do condomínio a gerir melhor e tomar ações com base nos dados para melhorar os processos, segurança e bem estar dos condôminos. Para as análises utilizamos os conhecimentos obtidos sobre PBI e python utilizando os métodos de soma, séries temporais, estacionariedade, autocorrelação, modelo Arima e outliers consolidando nossos conhecimentos de análise de dados.

Utilizamos também os conhecimentos em tecnologias cloud com o banco de dados NoSql da Azure, onde criamos a massa de dados para análise.

Foi uma experiência desafiadora para todo o grupo, onde foram necessários encontros, conversas e pesquisas para chegar ao resultado final.