

DESAFIO

DATA SCIENCE

elogroup▶



DESAFIO: EXPANSÃO DE UMA REDE DE LABORATÓRIOS

Análise de Negócio para a Empresa EloGroup

Dr Dheiver Francisco Santos
Cientista de Dados Sênior

Análise Exploratória

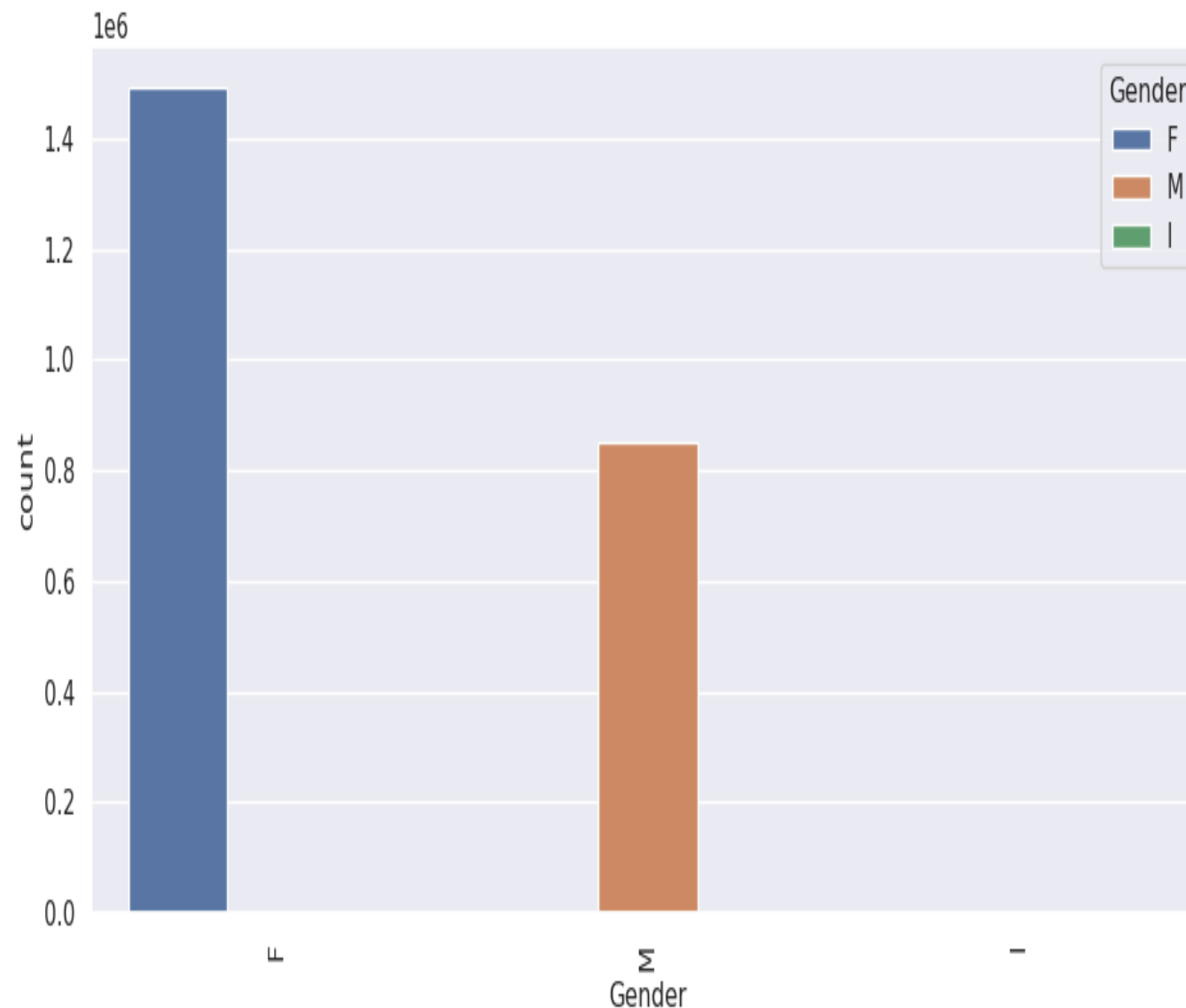
Análise quantitativa da distribuição de gênero

Analisando o Dataframe , observou-se três tipologias em relação ao gênero (**Feminino, Masculino e Indefinido**).

Você tem uma distribuição de **63%** do sexo feminino , **36%** do sexo Masculino, e uma porção ínfima de pessoas que marcaram como Indefinido.

Percentual:

F >> 0.636994 ; M >> 0.363002 ; I >> 0.000004



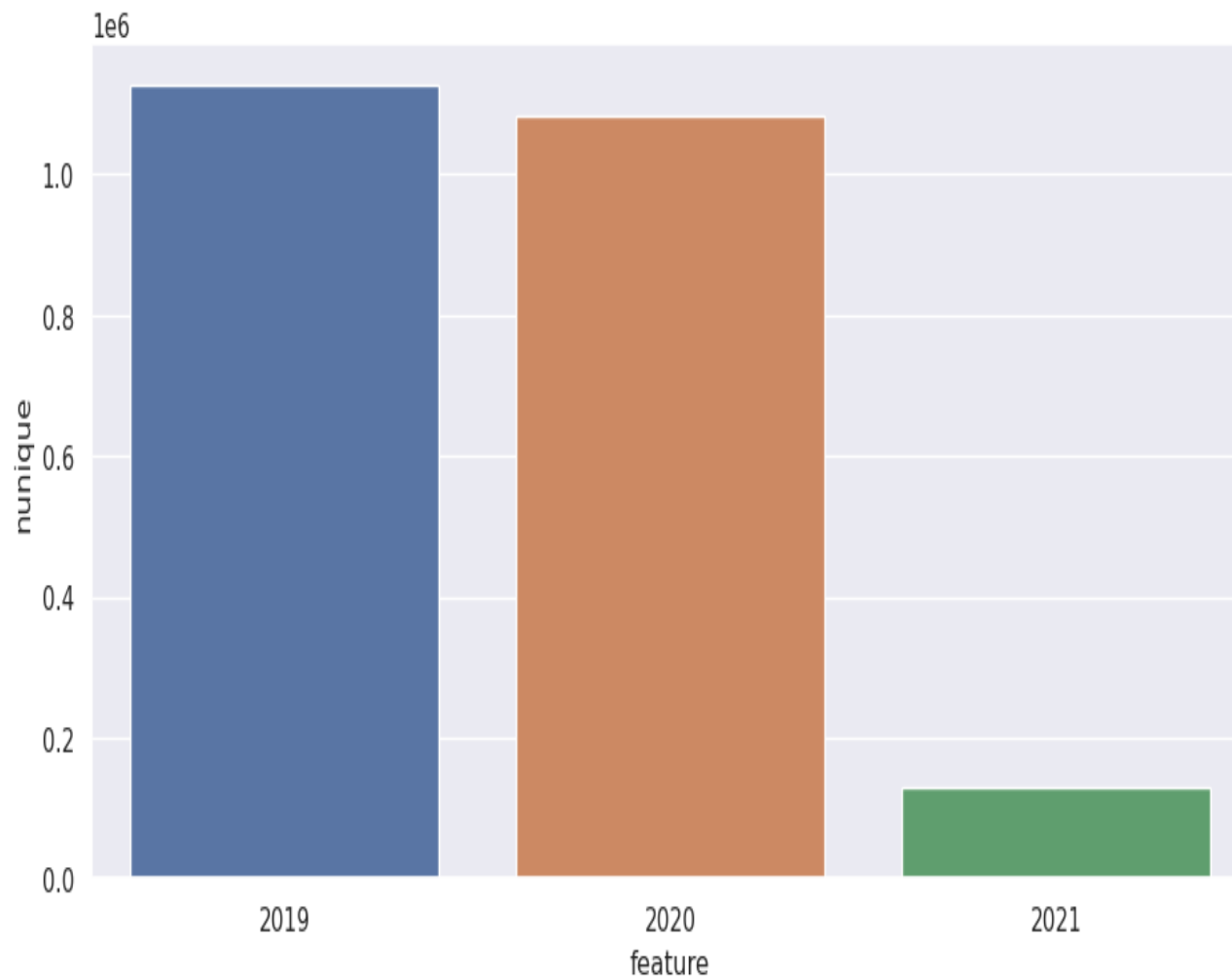
Análise quantitativa da distribuição do número de exames por ano

Analisando o Dataframe , observou-se três tipologias em relação ao ano (**2019, 2020 e 2021**).

Você tem uma distribuição de **48% no ano de 2019, 46 % no ano de 2020, e próximo a 5,5% no ano de 2021**. Sendo o ano de 2021 o ano de menor demanda, ou seja, no ano de 2019 e 2020 ocorreram mais exames que 2021.

Percentual:

```
2019>> 0.482148; 2020 >> 0.462226; 2021  
>> 0.055626
```



Análise quantitativa da distribuição do laboratório em que o paciente fez o exame clínico

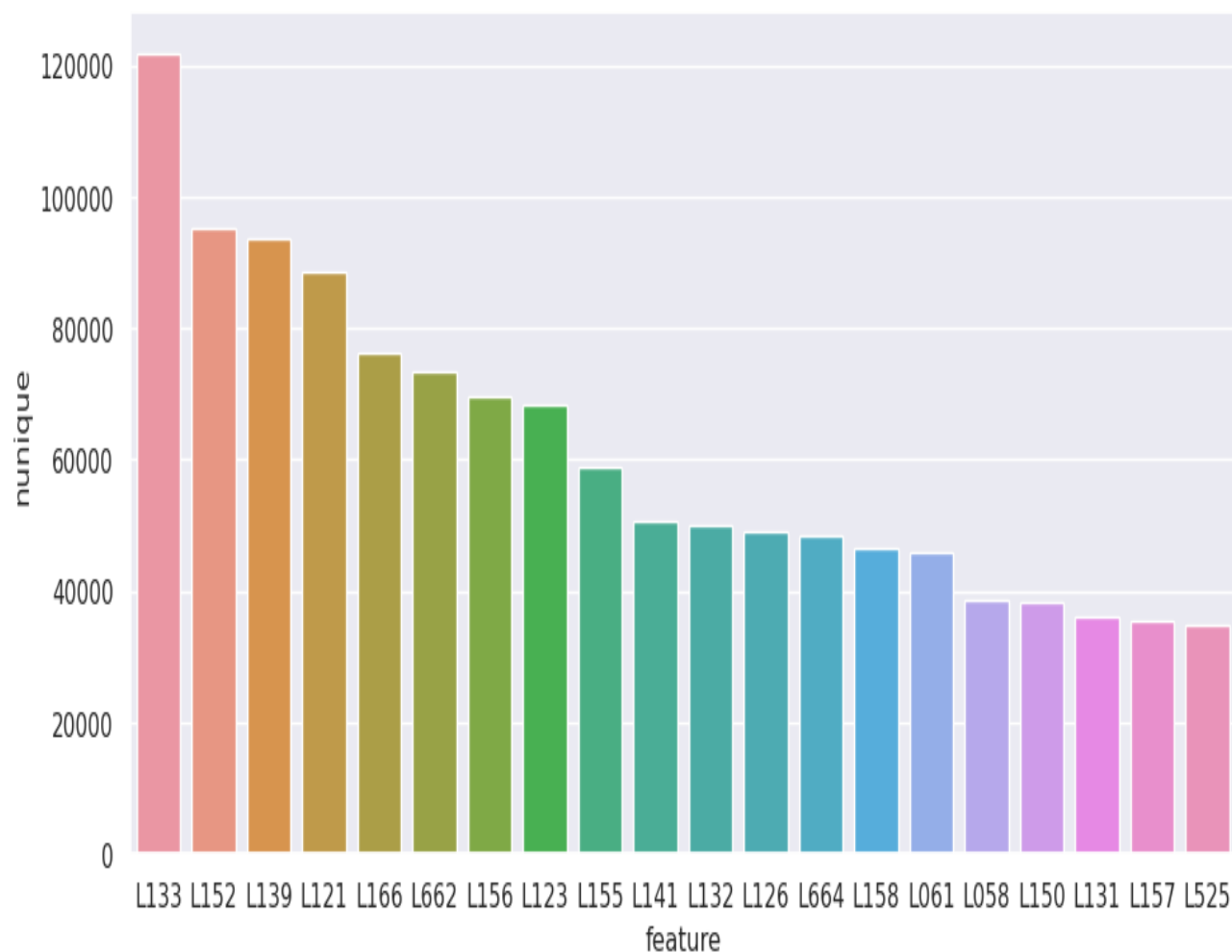
Analisando o Dataframe , observou-se a presença de 117 laboratórios de exames clínicos.

Você pode observar os principais laboratórios onde são realizados mais exames clínicos. Podemos destacar os laboratórios com ID **L133, L152, L139, L121, L166**.

Será que a região desses laboratório demandam novas unidades? É possível saber disso fazendo um forecast

Quantidade:

L133 >>121886; L152>>95255; L139>>93495; L121 >>88430; L166 >>76159

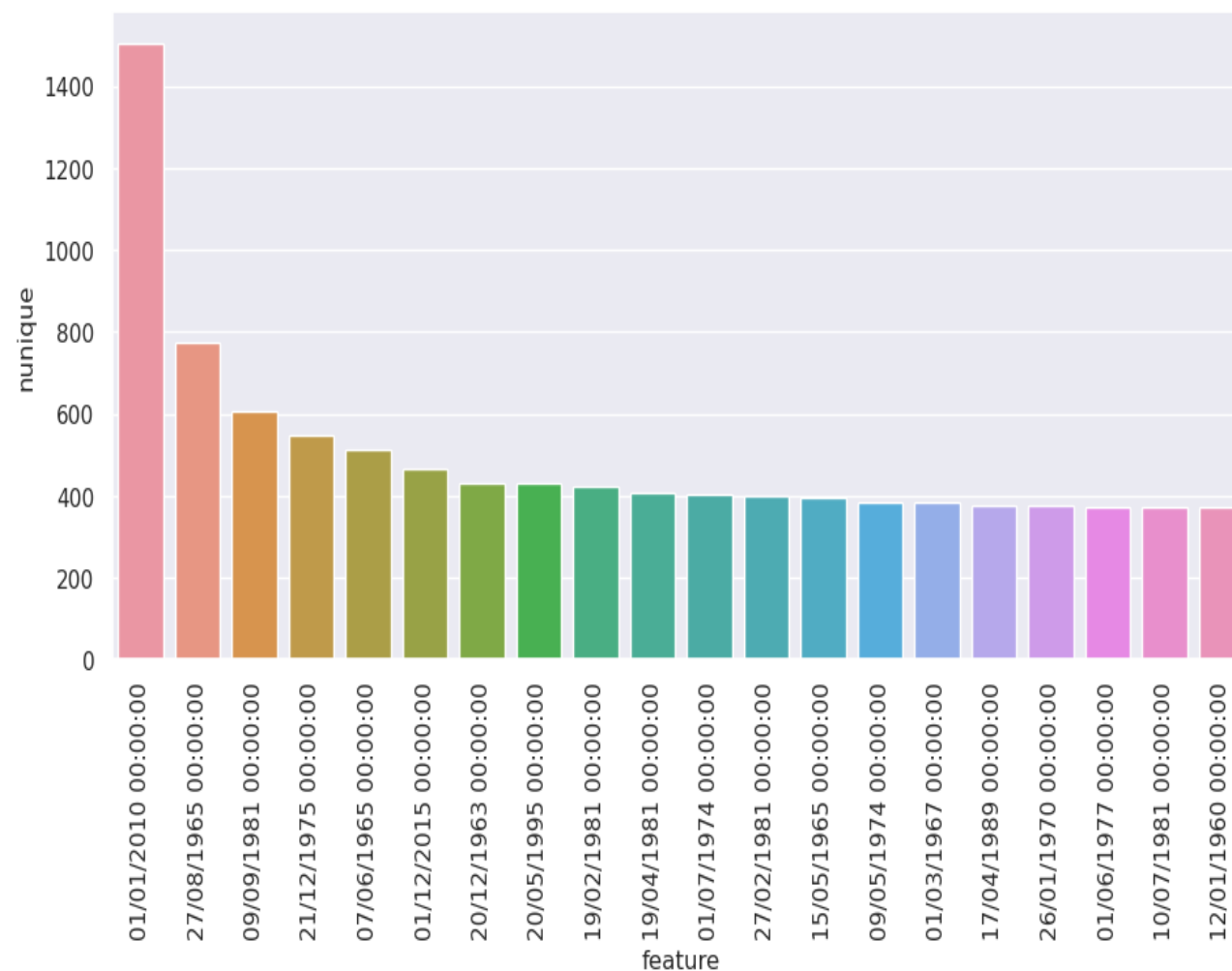


Análise quantitativa da distribuição do dia de aniversário dos clientes

Analisando o Dataframe , observou-se a os meses com maior recorrência de datas de aniversário dos clientes.

Você pode observar os principais aniversários são realizados nos meses de **janeiro, agosto, setembro e dezembro**.

“Aniversário do cliente: oportunidade de fidelização ... para gerar mais vendas, mas também para começar um relacionamento de fidelização. Esses meses poderiam ser usados para políticas de fidelização de clientes “



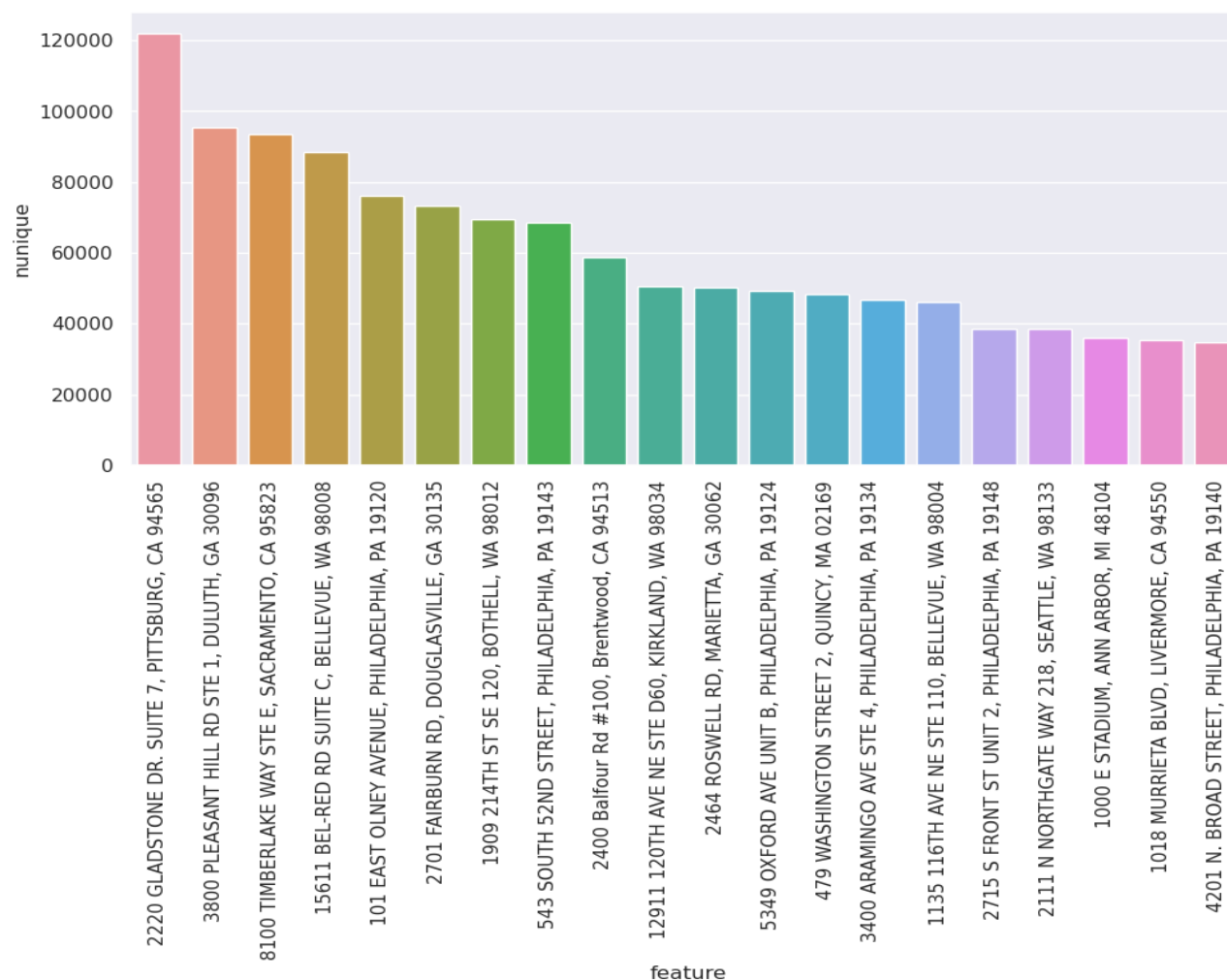
Análise quantitativa da distribuição dos principais endereços onde ocorrem exames clínicos

Analisando o Dataframe , observou-se a presença de 117 laboratórios de exames clínicos.

Você pode observar os principais laboratórios onde são realizados mais exames clínicos. Podemos destacar o laboratório de nome : **2220 GLADSTONE DR. SUITE 7, PITTSBURG, CA 94565** com um total de **121886** registros.

Esse Lab possui ID L133

Será que a região desses laboratório demandam novas unidades? É possível saber disso fazendo um forecast

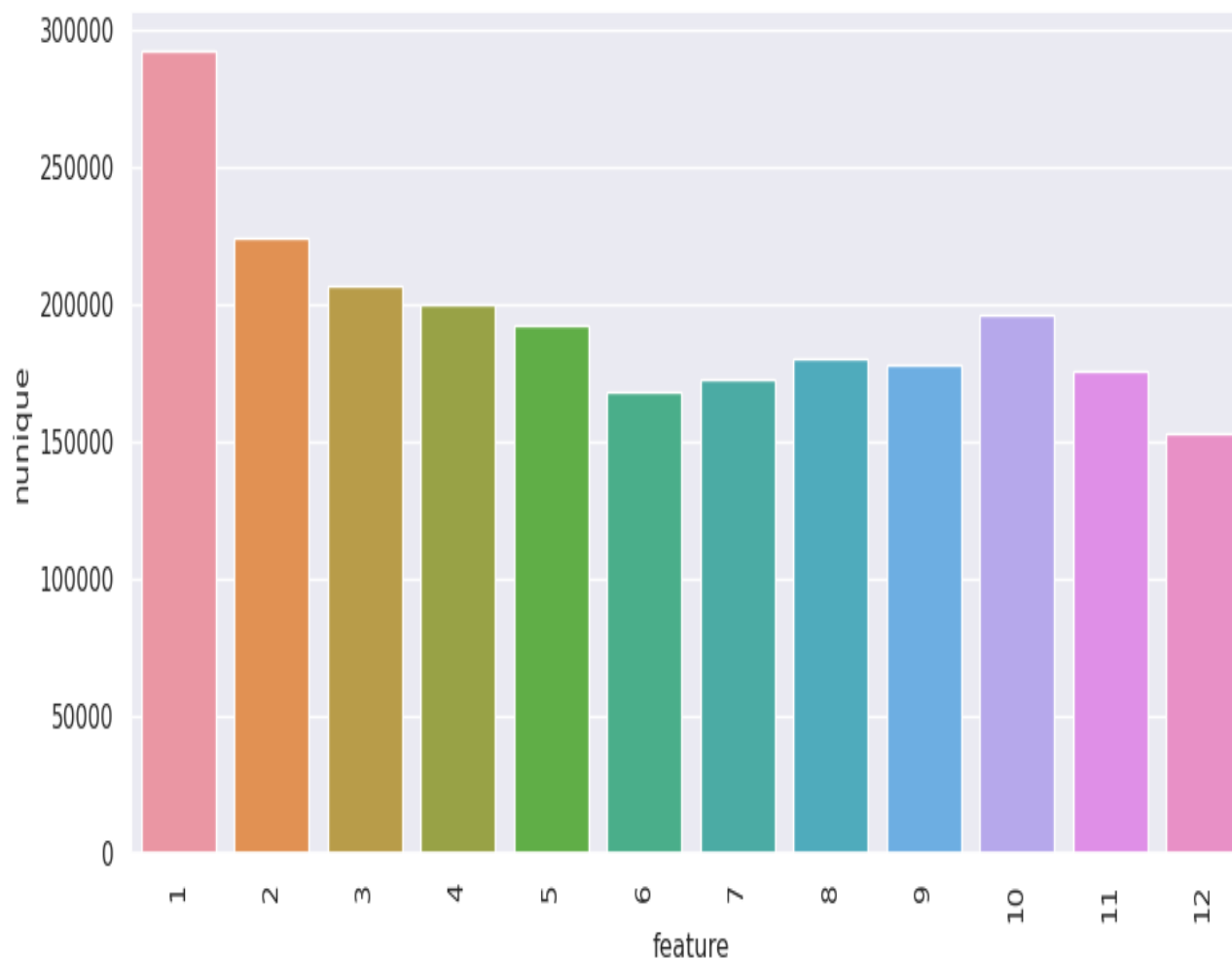


Análise quantitativa da distribuição dos principais meses de ocorrência do número da realização de exames clínicos

Analizando o Dataframe em relação a Análise quantitativa da distribuição dos principais meses de ocorrência do número da realização de exames clínicos, foi percebido um crescimento acima da média no mês de janeiro e fevereiro.

Ou seja, os meses de **janeiro e fevereiro** são os meses de alta demanda de exames médicos requerendo um plano de ação especial para esse período (*por exemplo: aumento de contratações temporário para atender a demanda*).

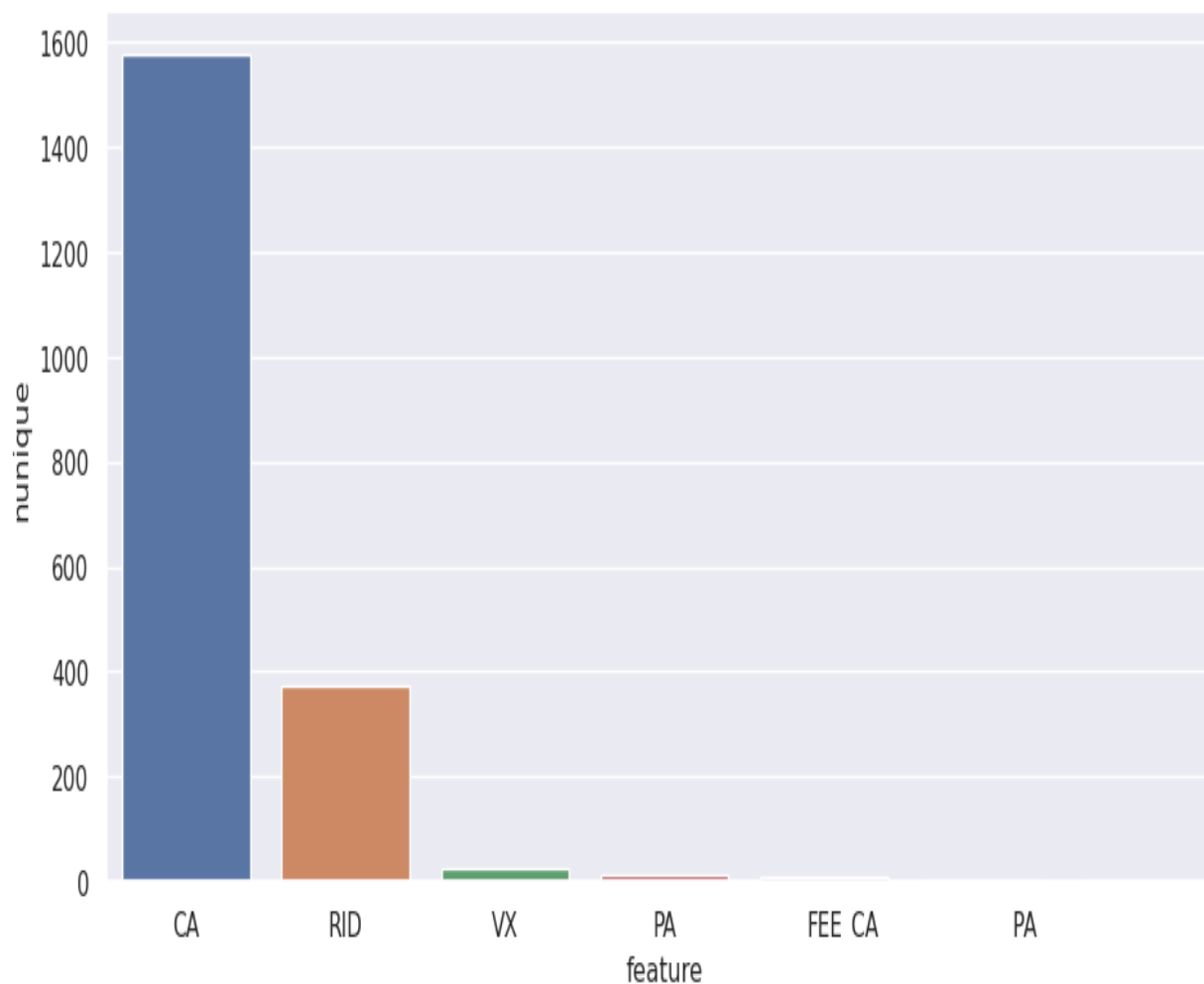
EDA – Análise Exploratória



Análise quantitativa da distribuição das principais categorias de exames clínicos realizados

Analisando o Dataframe em relação a Análise quantitativa da distribuição das principais categorias de exames clínicos realizados, foi percebido dois principais grupamentos em relação a essas categorias. **As categorias de destaque são a CA e RD, com respectivamente 78 % e 18%.**

“O que reflete a necessidade de um direcionamento orçamentário para esses grupos ao citarmos expansão de atividades.”



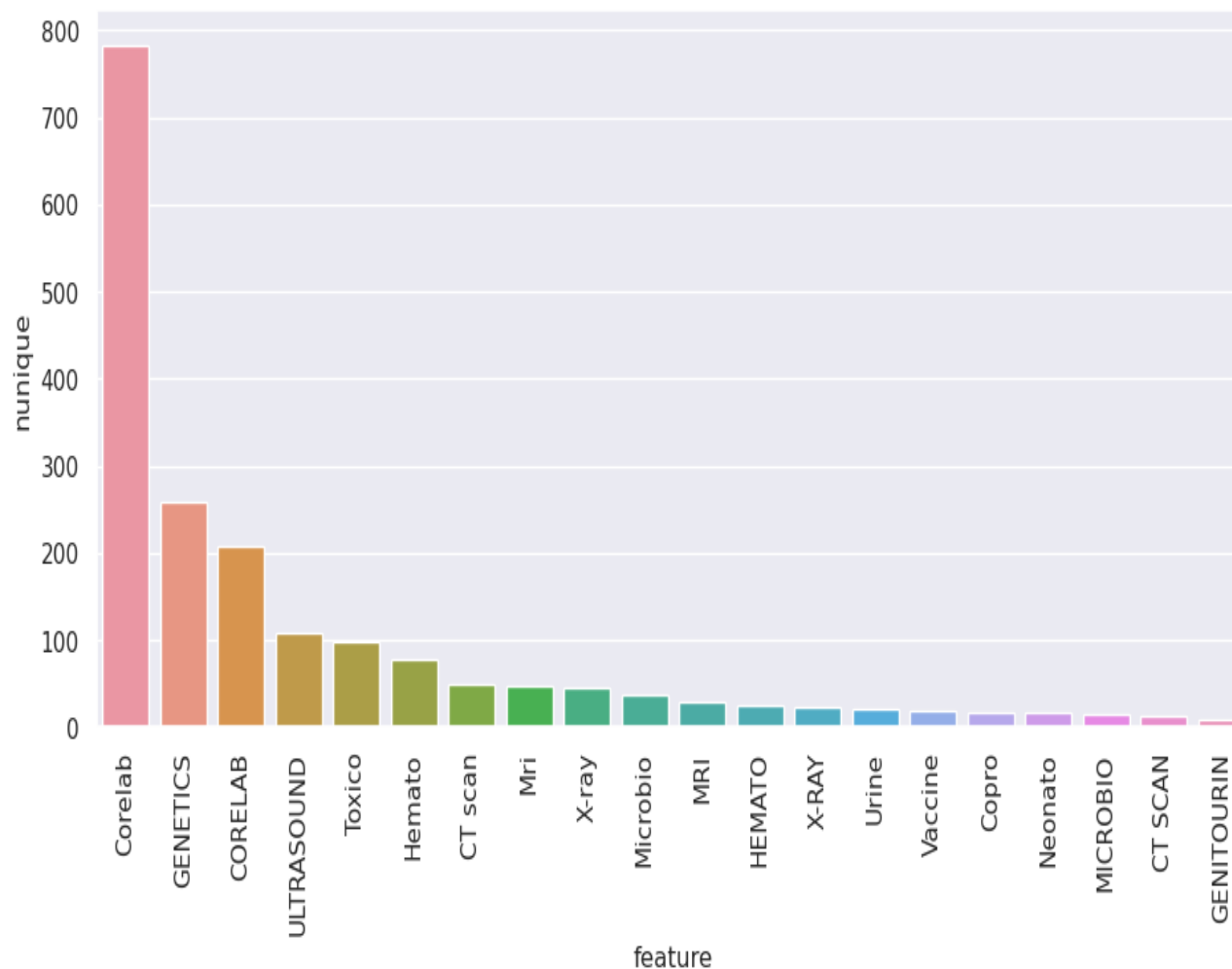
Análise quantitativa da distribuição das principais famílias de exames clínicos realizados

Analisando o Dataframe em relação a Análise quantitativa da distribuição das principais famílias de exames clínicos realizados, foi percebido 7 principais grupamentos em relação a essas famílias.

As famílias em destaque são **Corelab, GENETICS, CORELAB, ULTRASOUND, Toxico Hemato, CT scan**.

“O que reflete a necessidade de um direcionamento orçamentário para essas famílias ao citarmos expansão de atividades.”

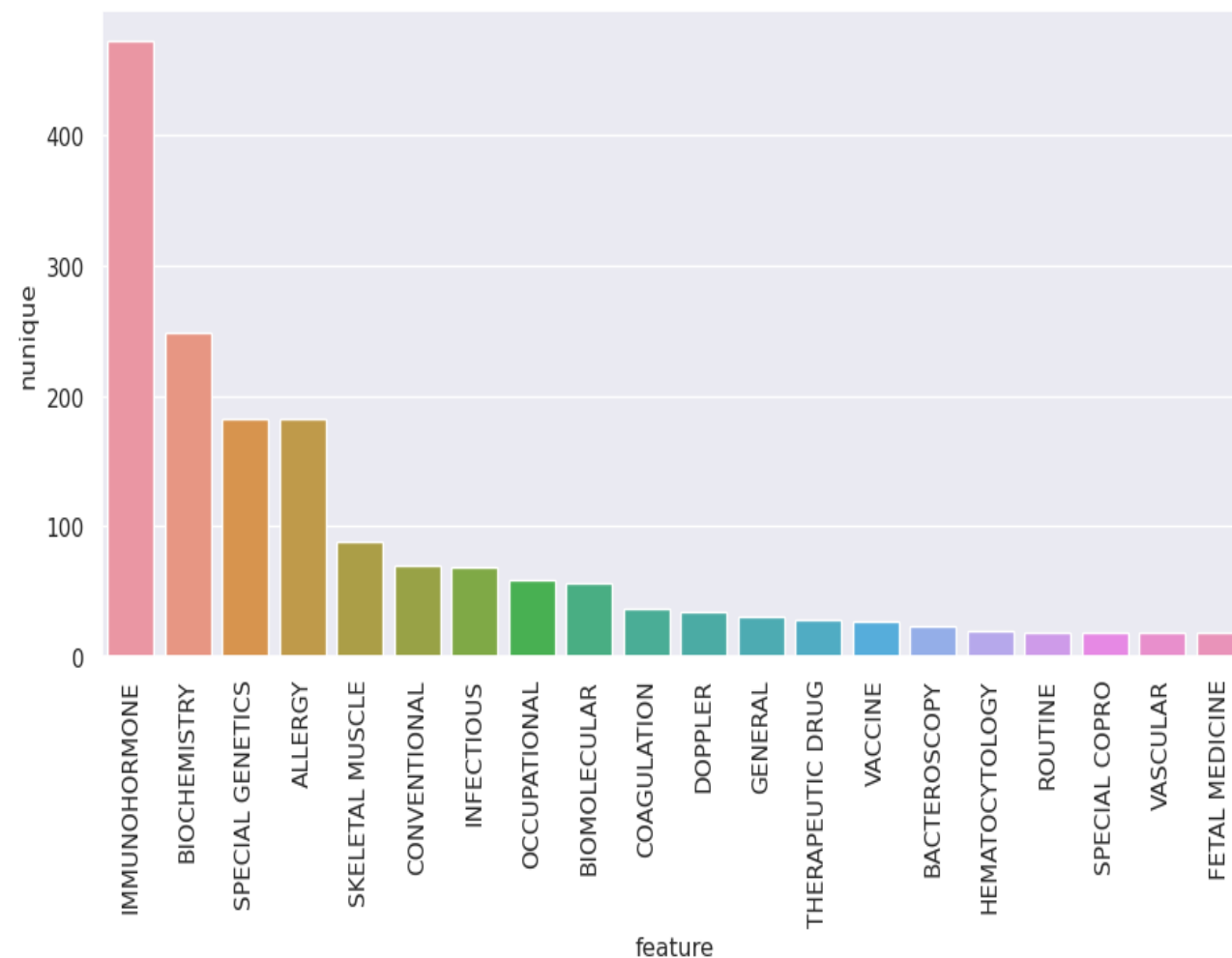
EDA – Análise Exploratória



Análise quantitativa da distribuição das principais especialidades de exames clínicos realizados

Analisando o Dataframe em relação a Análise quantitativa da distribuição das principais especialidades de exames clínicos realizados, foi percebido 5 principais grupamentos em relação a essas especialidades. As especialidades em destaque são:

IMMUNOHORMONE (equivalente a 23%), **BIOCHEMISTRY** (equivalente a 12%), **SPECIAL GENETICS** (equivalente a 9%), **ALLERGY** (equivalente a 9%), **SKELETAL MUSCLE** (equivalente a 4%), “O que reflete a necessidade de um direcionamento orçamentário para essas especialidades ao citarmos expansão de atividades.”

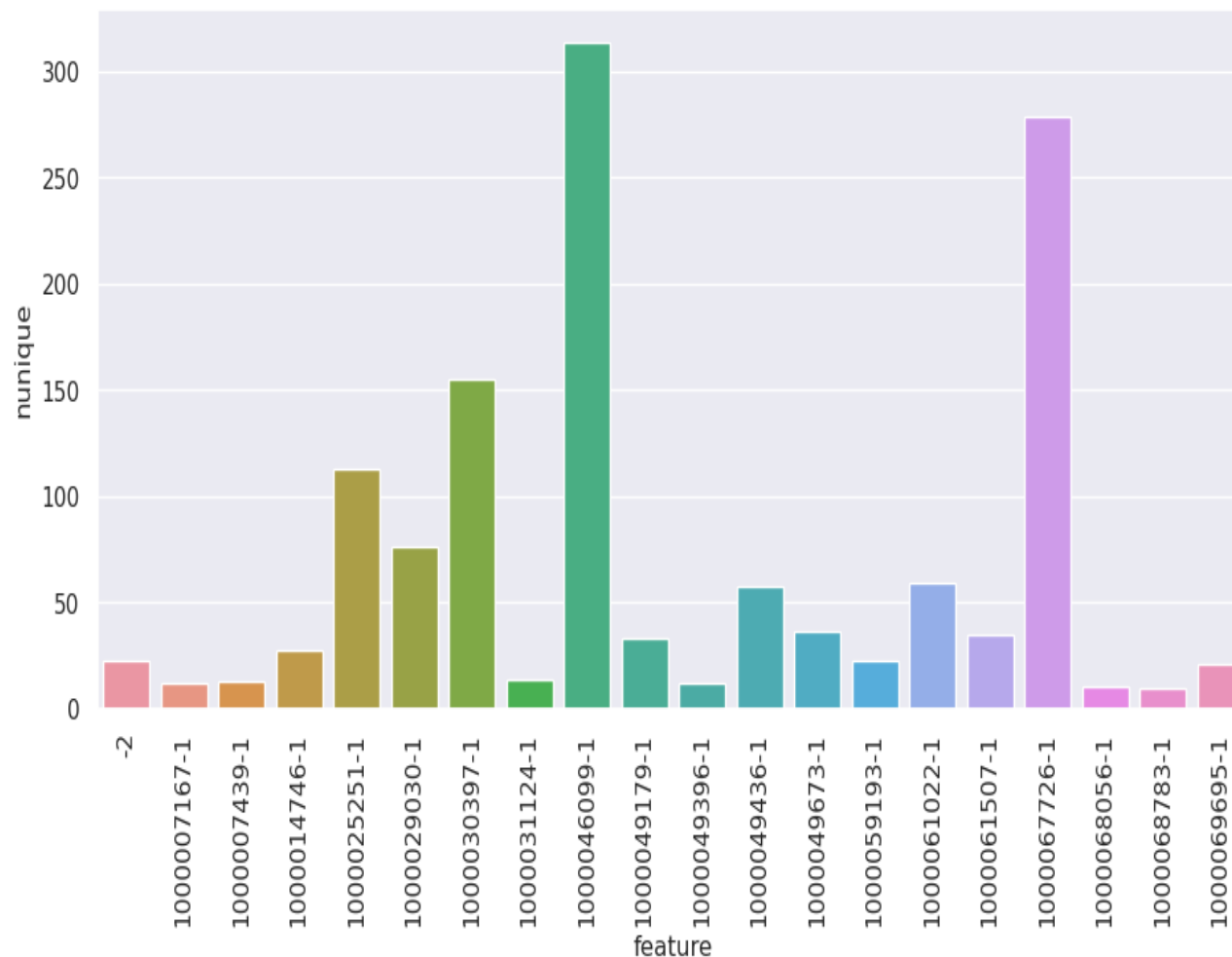


Análise quantitativa da distribuição das principais clientes em relação à exames clínicos realizados

Analisando o Dataframe em relação a Análise quantitativa da distribuição das principais clientes em relação à exames clínicos realizados, foi percebido um grupo especial de clientes que mais deixam recursos na organização. **São eles final :51, 30 , 97, 99, 26 .**

“A descoberta desses clientes é importante para que ações de fidelização dos mesmos posam ser executadas como prioritária.”

EDA – Análise Exploratória

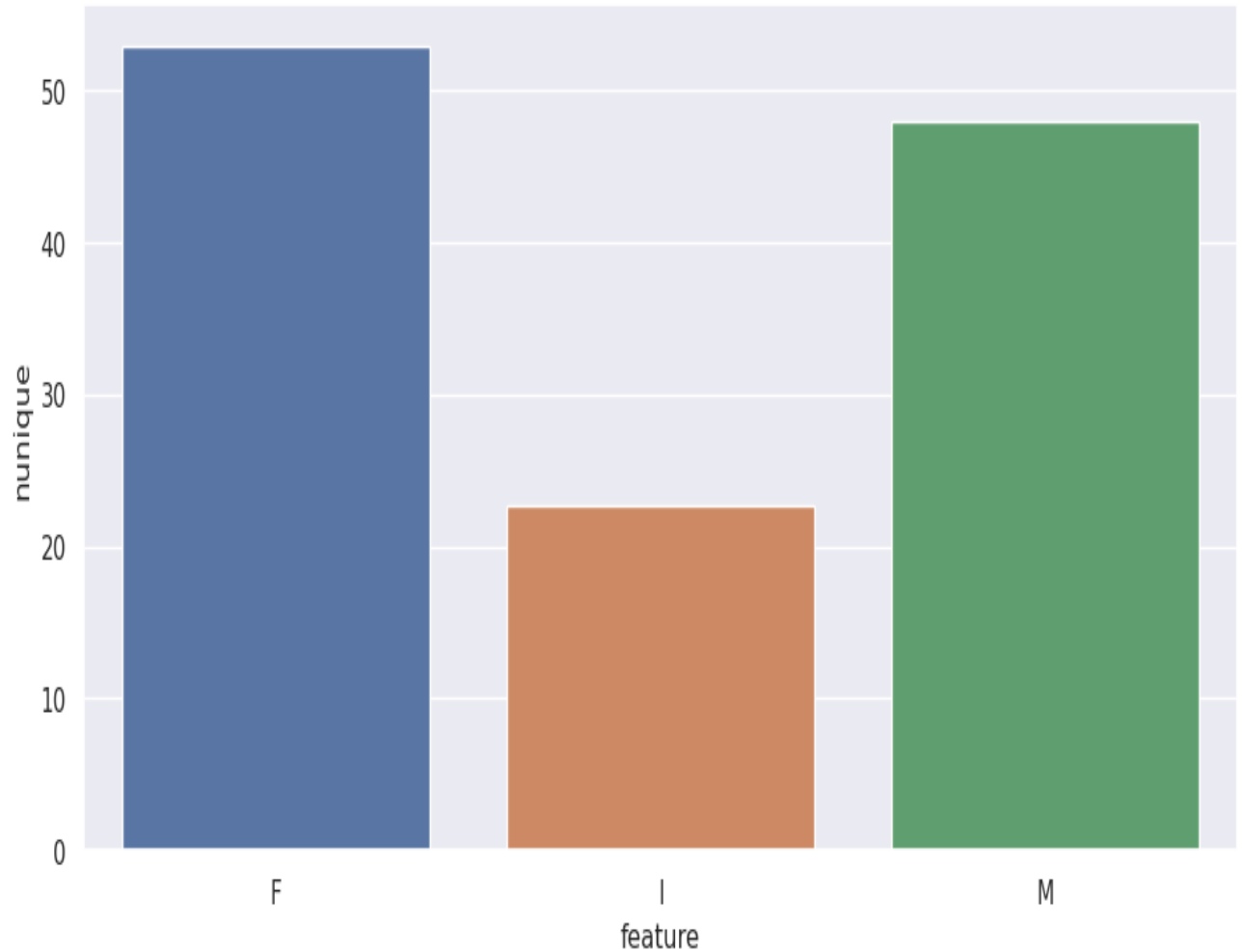


Análise quantitativa da distribuição do Ticket médio em relação ao gênero

Analisando o Dataframe em relação a **Análise quantitativa da distribuição do Ticket médio em relação ao gênero**, foi percebido a distribuição entre masculino, feminino e indefinido.

Mulheres gastam em média (F-52.936704), indefinidos gastam em média (I-22.666667), e Homens gastam em média (M -47.8837550).

Esse tipo de análise ajuda no processo de fidelização e definição de políticas de combos e descontos.

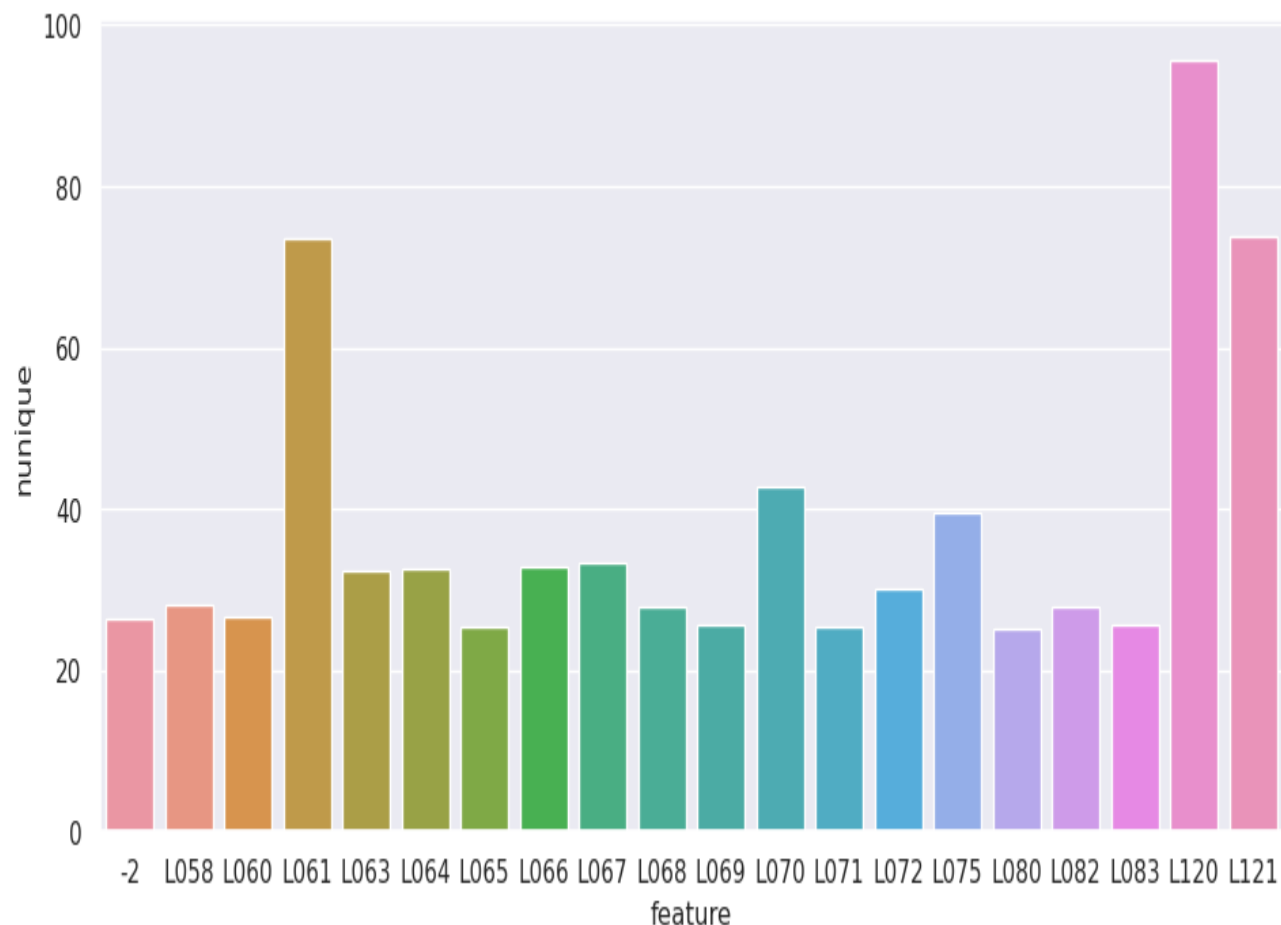


Análise quantitativa da distribuição do Ticket médio em relação ao ID do laboratório

EDA – Análise Exploratória

Analisando o Dataframe em relação a **Análise quantitativa da distribuição do Ticket médio em relação ao ID do laboratório**, foi percebido maior distribuição para **L61, L120, L121**.

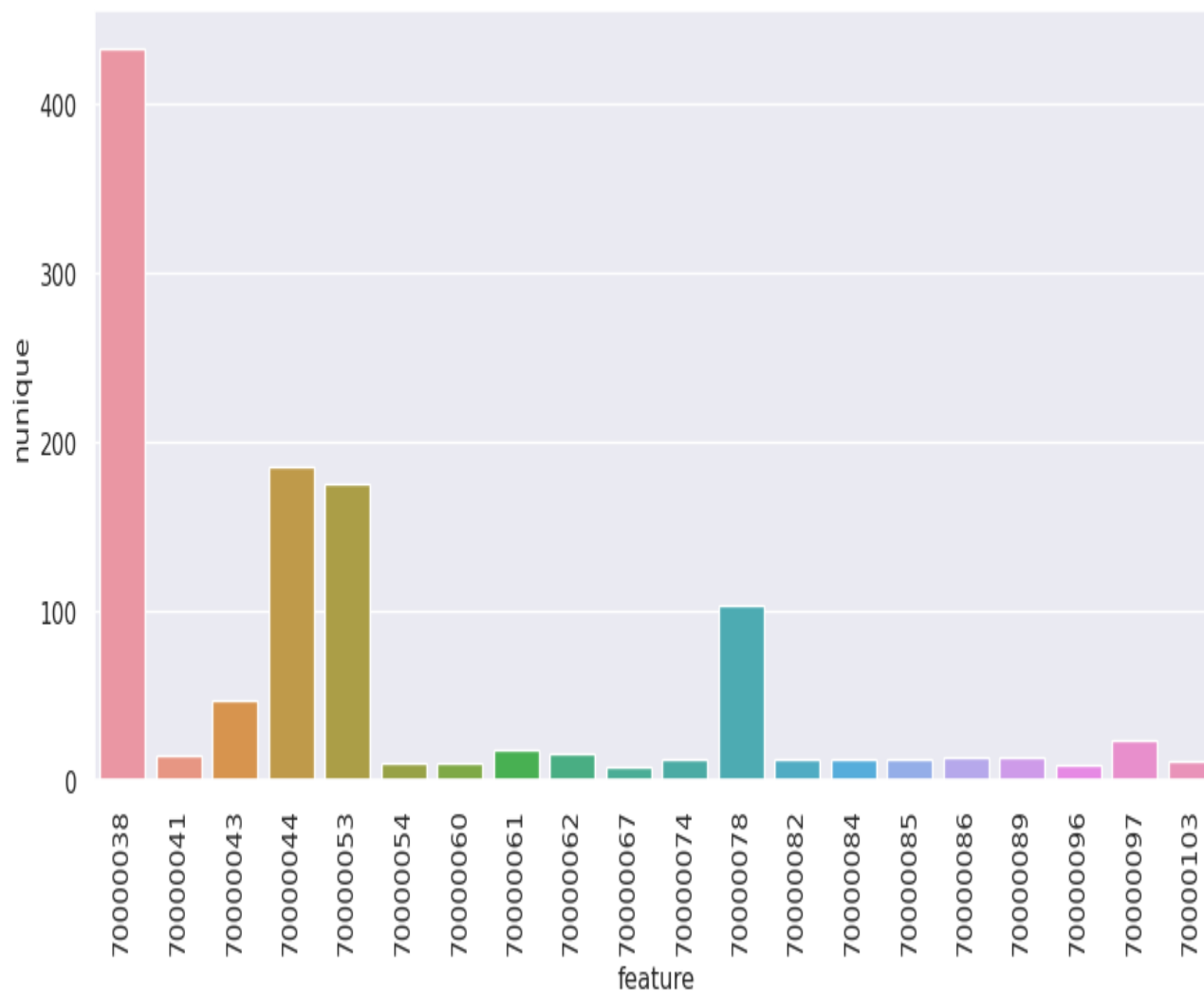
*Apesar desses laboratórios possuírem maior ticket médio, somente **L121** possui um excelente quantitativo de exames realizados. Os laboratórios que mais realizam exames são : ID **L133, L152, L139, L121, L166**.*



Análise quantitativa da distribuição do Ticket médio em relação ao código de exame de laboratório

Analisando o Dataframe em relação a **Análise quantitativa da distribuição do Ticket médio em relação ao código de exame de laboratório**, foi percebido maior distribuição para final **38, 43, 53 e 78**.

Teoricamente esses ID de exames deixam na média mais recursos na instituição, o que corrobora para uma situação de priorização orçamentária no processo de expansão.



Análise quantitativa da distribuição do Ticket médio em relação ao endereço do laboratório

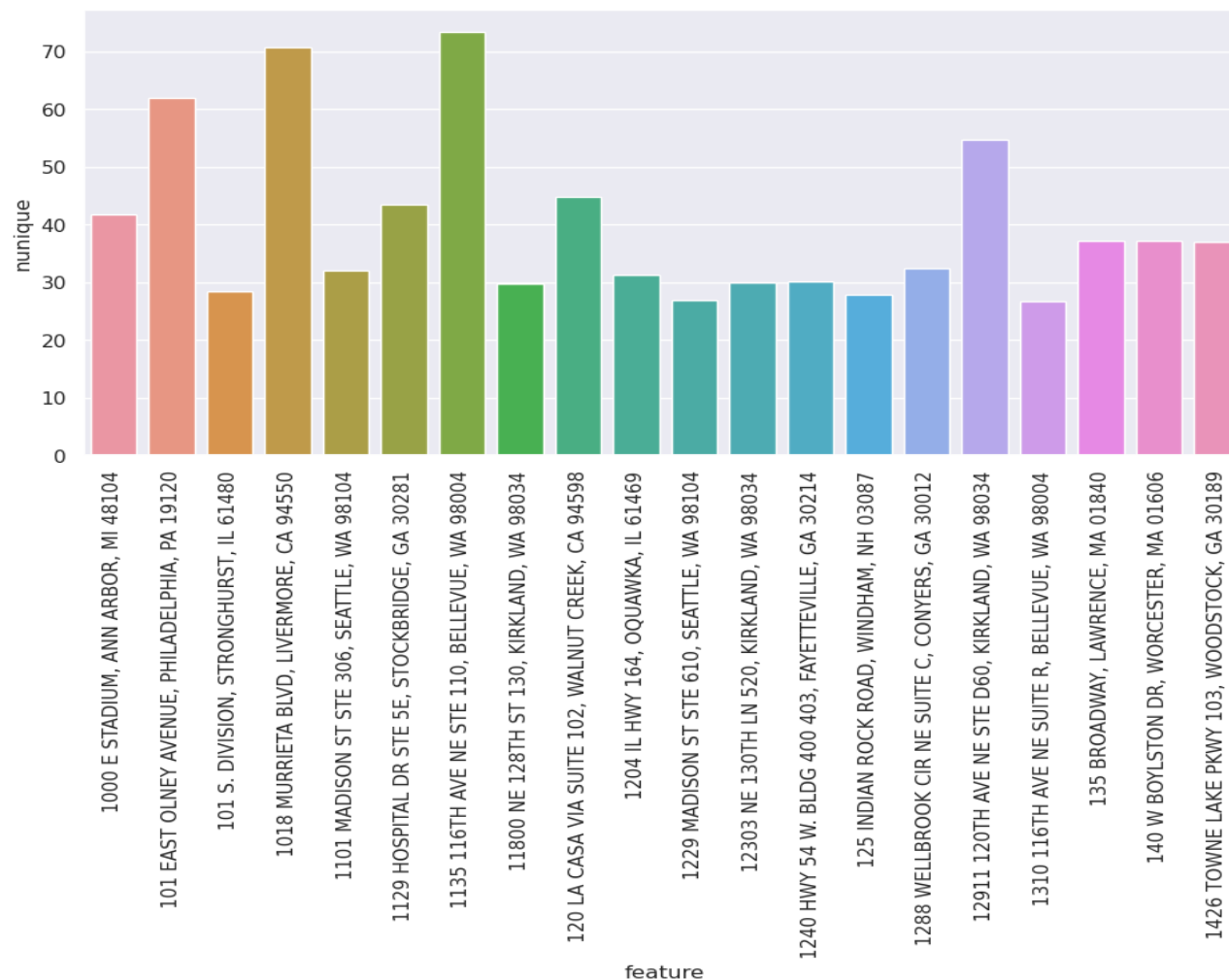
Analisando o Dataframe em relação a **Análise quantitativa da distribuição do Ticket médio em relação ao endereço do laboratório**, foi percebido maior distribuição para

1000 E STADIUM, ANN ARBOR, MI 48104 41.837607

101 EAST OLNEY AVENUE, PHILADELPHIA, PA 19120 61.989457

101 S. DIVISION, STRONGHURST, IL 61480 28.470506

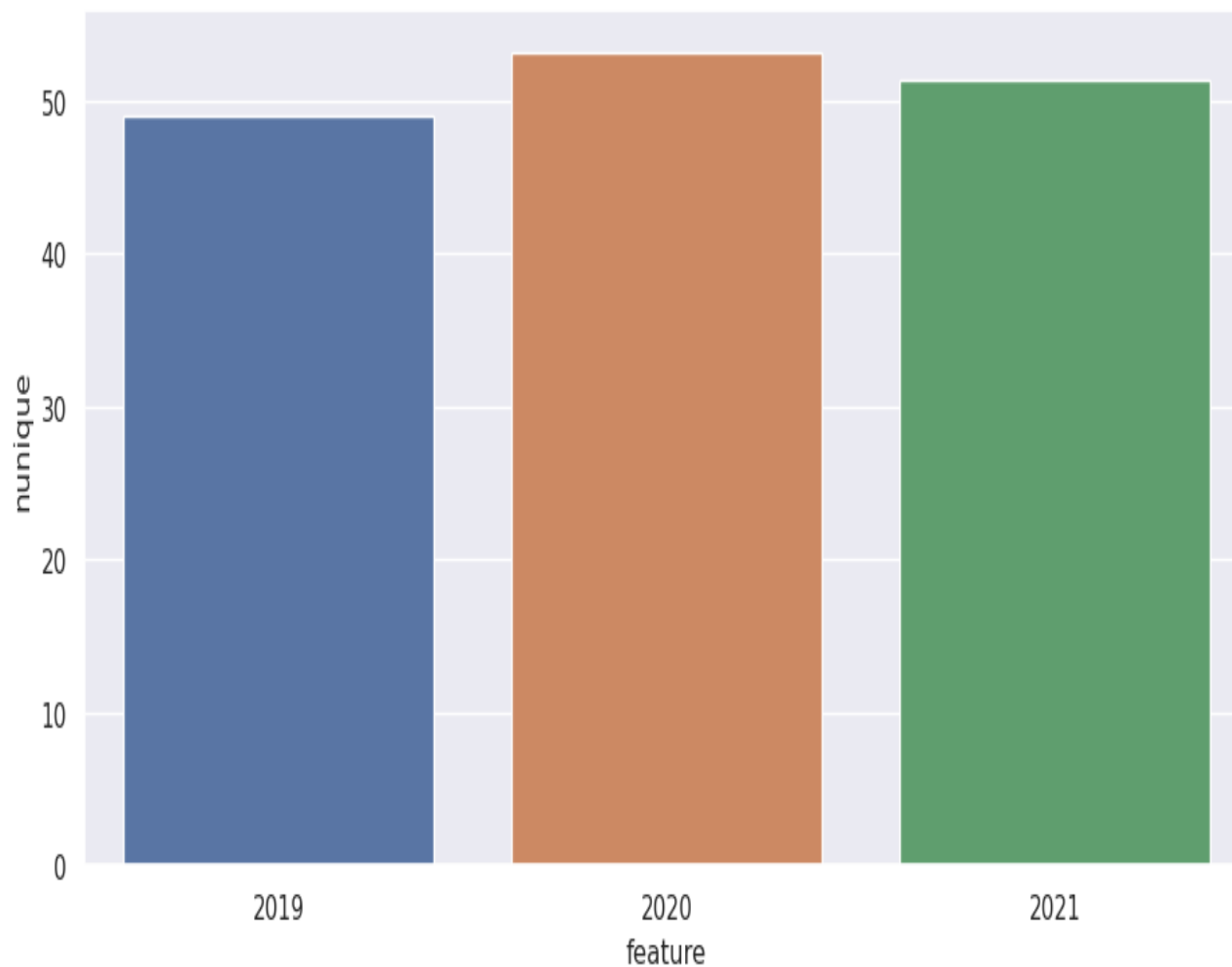
Atenção: Os principais endereços de maiores tickets não coincidem com os endereços que tem maiores números de exames. E isso é muito importante para tomada de decisão.



Análise quantitativa da distribuição do Ticket médio em relação ao ano de ocorrência do exame

Analisando o Dataframe em relação a **Análise quantitativa da distribuição do Ticket médio em relação ao ano de ocorrência do exame**, foi percebido distribuição equivalentes para os anos de 2019, 2020 e 2021.

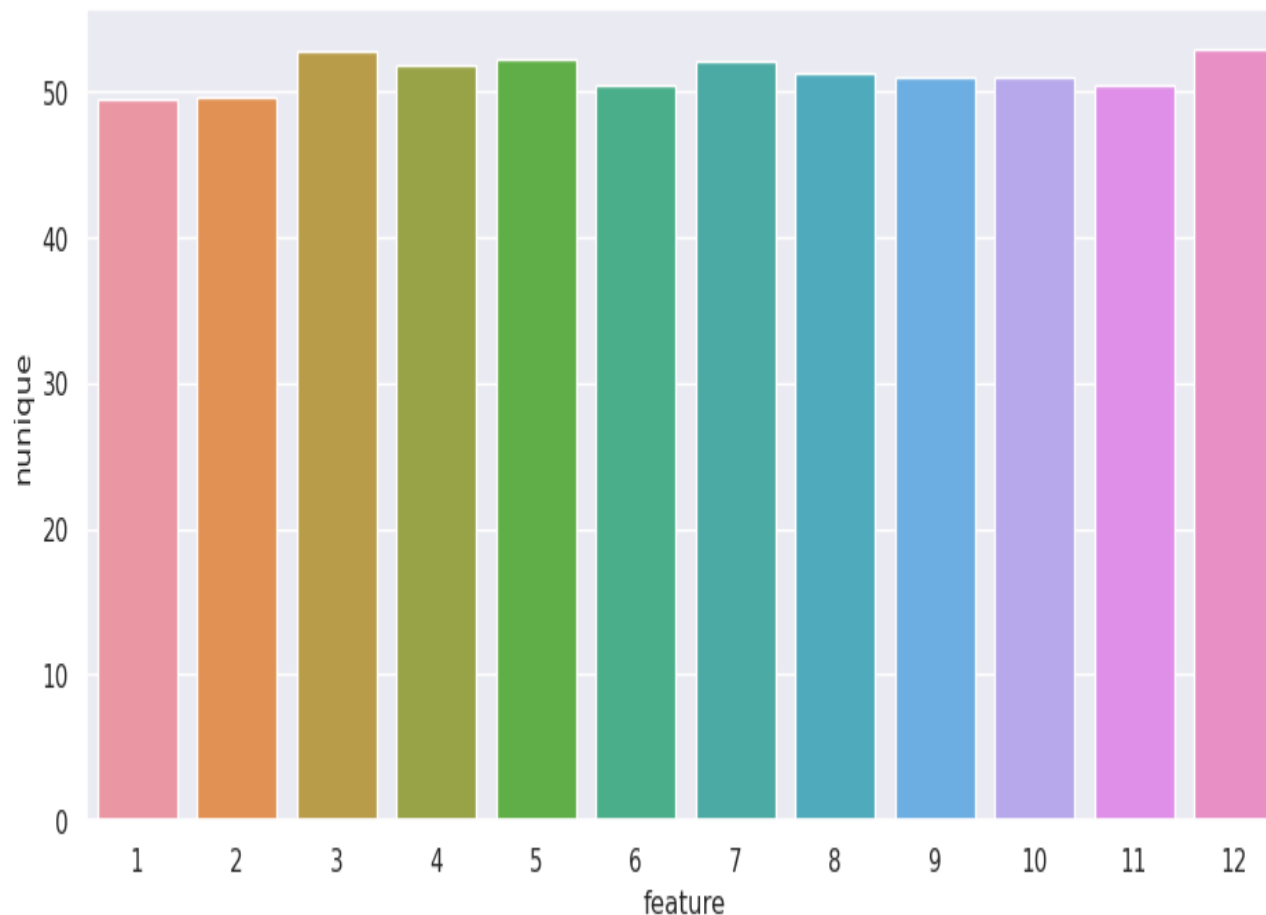
Isso mostra sustentação da base de clientes, bem como consistência das políticas de fidelização da organização.



Análise quantitativa da distribuição do Ticket médio em relação aos meses de realização do exame

Analizando o Dataframe em relação a **Análise quantitativa da distribuição do Ticket médio em relação aos meses de realização do exame**, foi percebido distribuição equivalentes para todos os meses durante os três anos de análise.

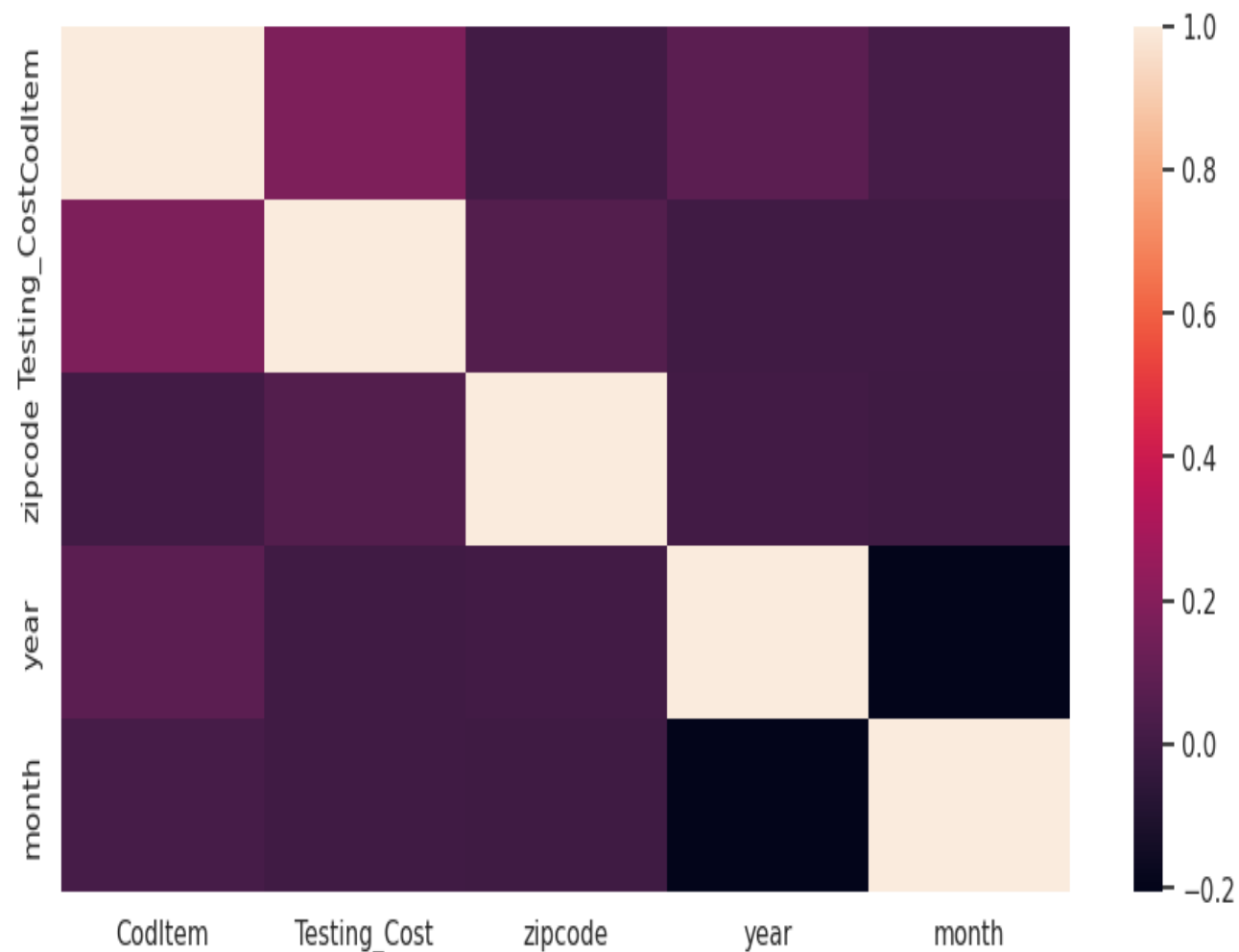
Isso mostra sustentação da base de clientes, bem como consistência das políticas de fidelização da organização.



Análise de correlação de variáveis para os dados de laboratório

Analizando o Dataframe em relação a **Análise de correlação de variáveis para os dados de laboratório**, assim, áreas mais claras apresentam forte correlação de dados e áreas mais escuras representam baixa correlação de dados.

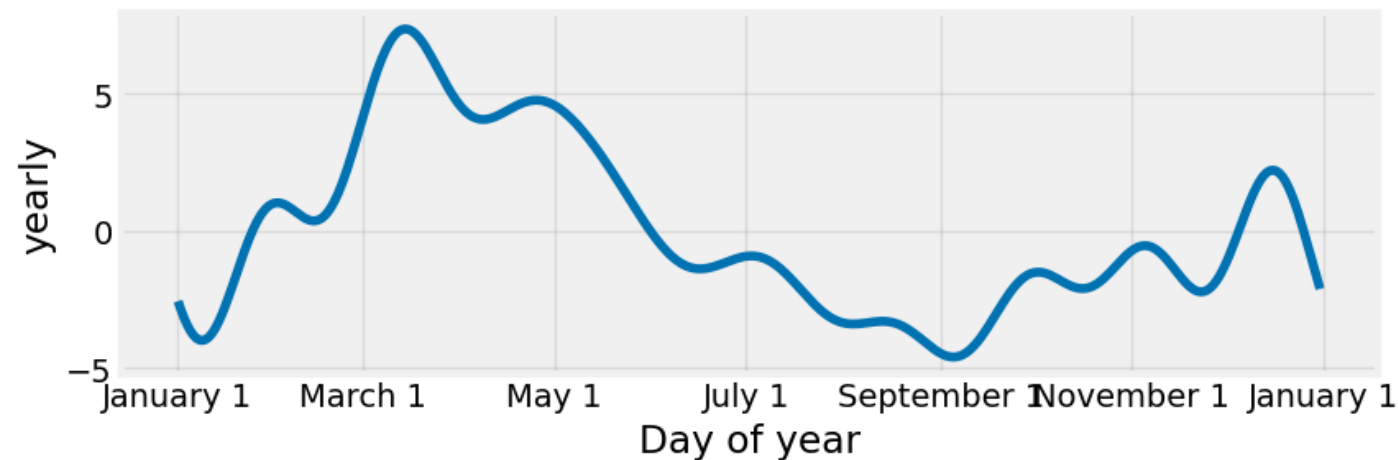
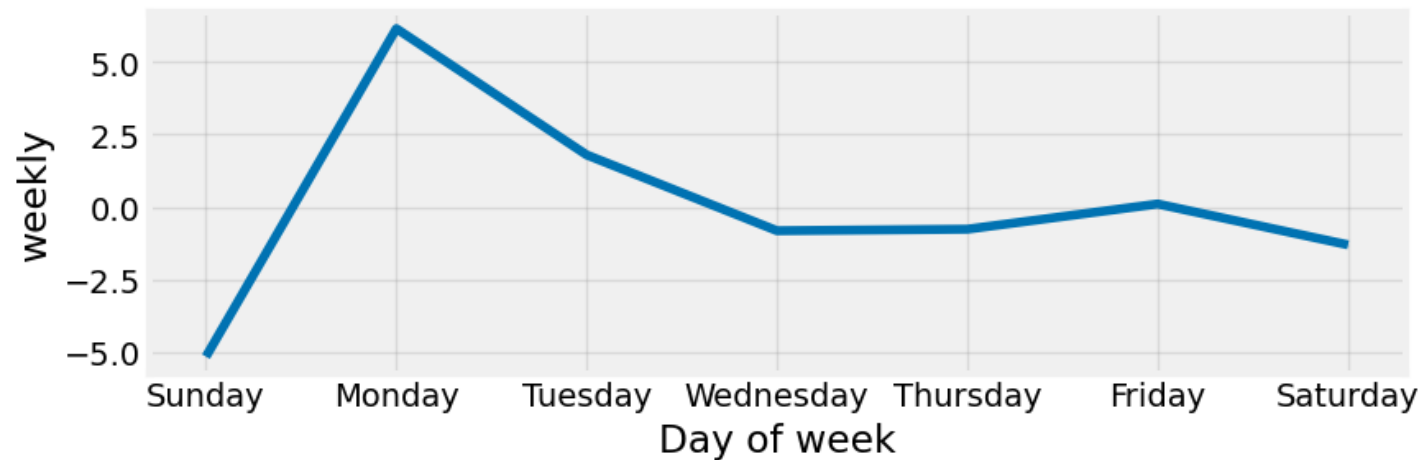
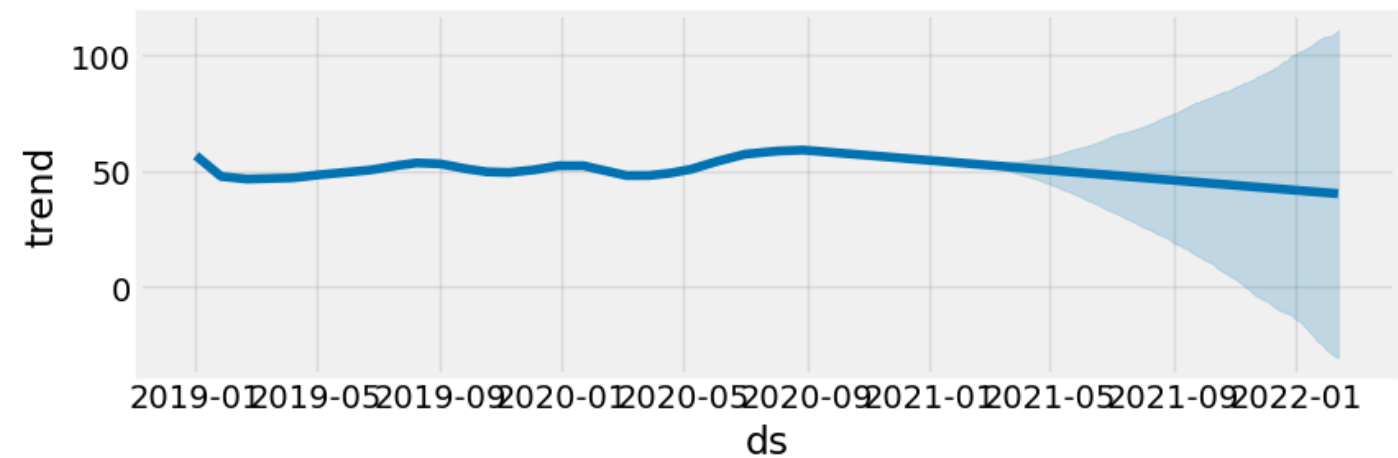
Por exemplo : (1) o código do item do exame tem boa correlação com o preço do exame, (2) o ano e o mês possui baixa correlação entre eles.



Forecast

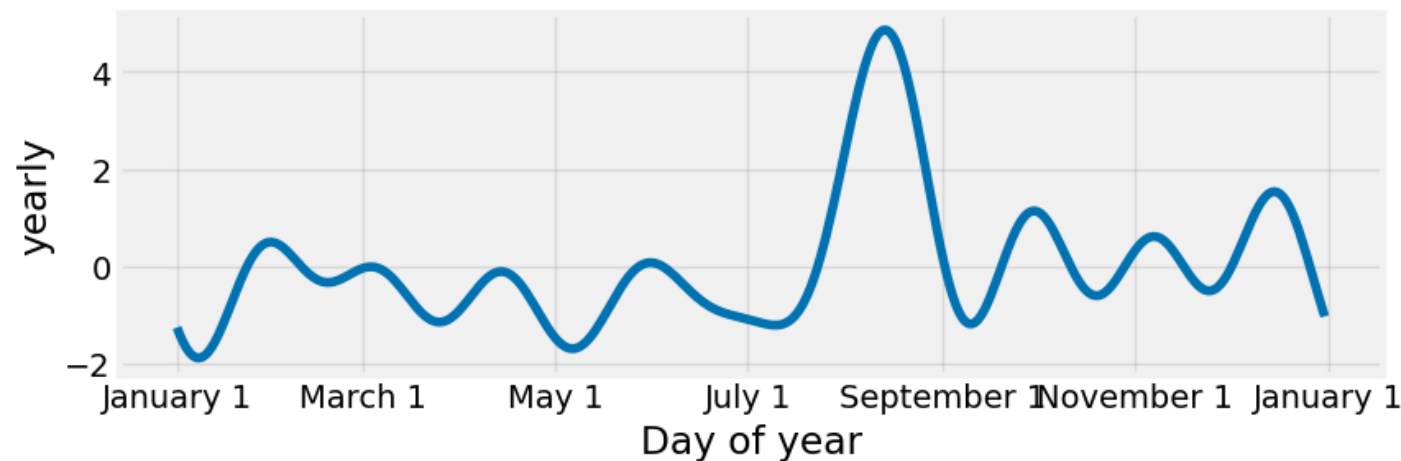
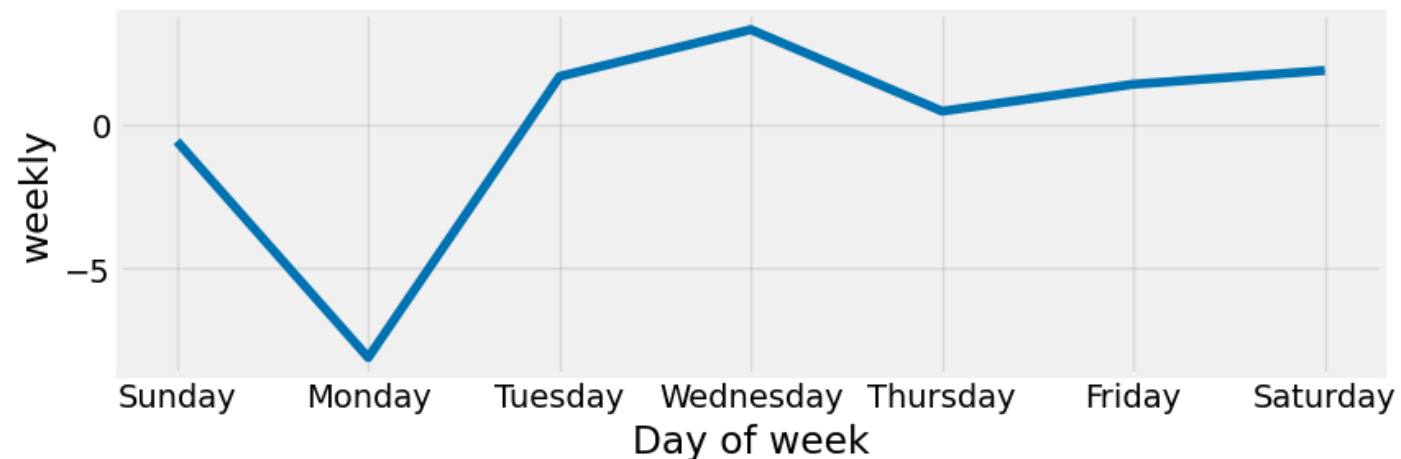
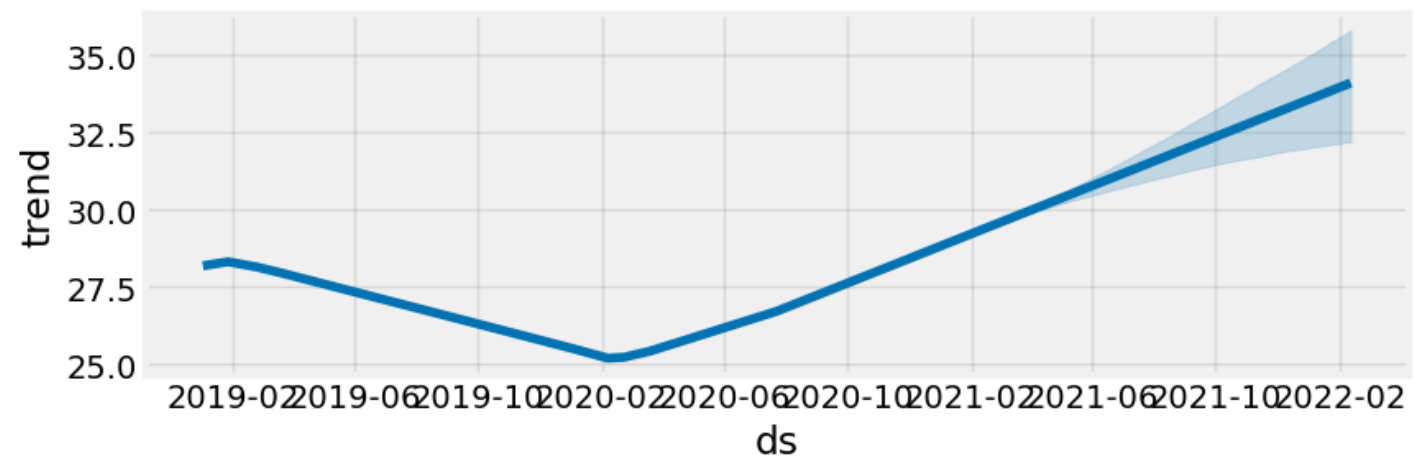
Previsão de preços de exames laboratoriais utilizando toda base histórica

Os resultados do Fbprophet mostrou uma direção indefinida de tendência em relação a crescimento e decaimento. Além disso, semanalmente mostrou uma tendência ascendente, então começou a se estabilizar, perdendo -2,00 de valor de mercado. O que mostra que o início da semana é bom para vendas. Por último, Anualmente, mostra uma tendência ascendente de janeiro a maio de 2022.



Previsão de preços de exames laboratoriais utilizando base histórica do laboratório L058

Os resultados do Fbprophet mostrou uma direção de crescimento em relação ao numero de vendas do laboratório L058. Além disso, semanalmente mostrou uma tendência ascendente, então começou a se estabilizar, Crescendo 5% de valor de mercado de segunda a sexta feira. Por último, Anualmente, mostra uma tendência de pico nos meses de julho a setembro de 2022. Esses resultados mostram claramente a necessidade de expansão dos serviços na região do laboratório L058.



Conclusão

O desafio foi realizado com sucesso ao ser apresentado uma análise exploratória e previsão de preços em relação ao comportamento dos diferentes laboratórios para que seja subsídio de tomada de decisão em relação ao processo de expansão da rede.