PROJETO II

Previsão de séries temporais

Trabalho realizado por:

- Ana Patrícia Silva - a22211661
- Joana Lopes - a22210868
- Tiago Vieira - a22210647

PROJETO II

BUSINESS UNDERSTANDING

SALES

- Store ID
- Product ID
- Vendas
- Stock
- Preço
- Receitas
- Promoções

PRODUCT

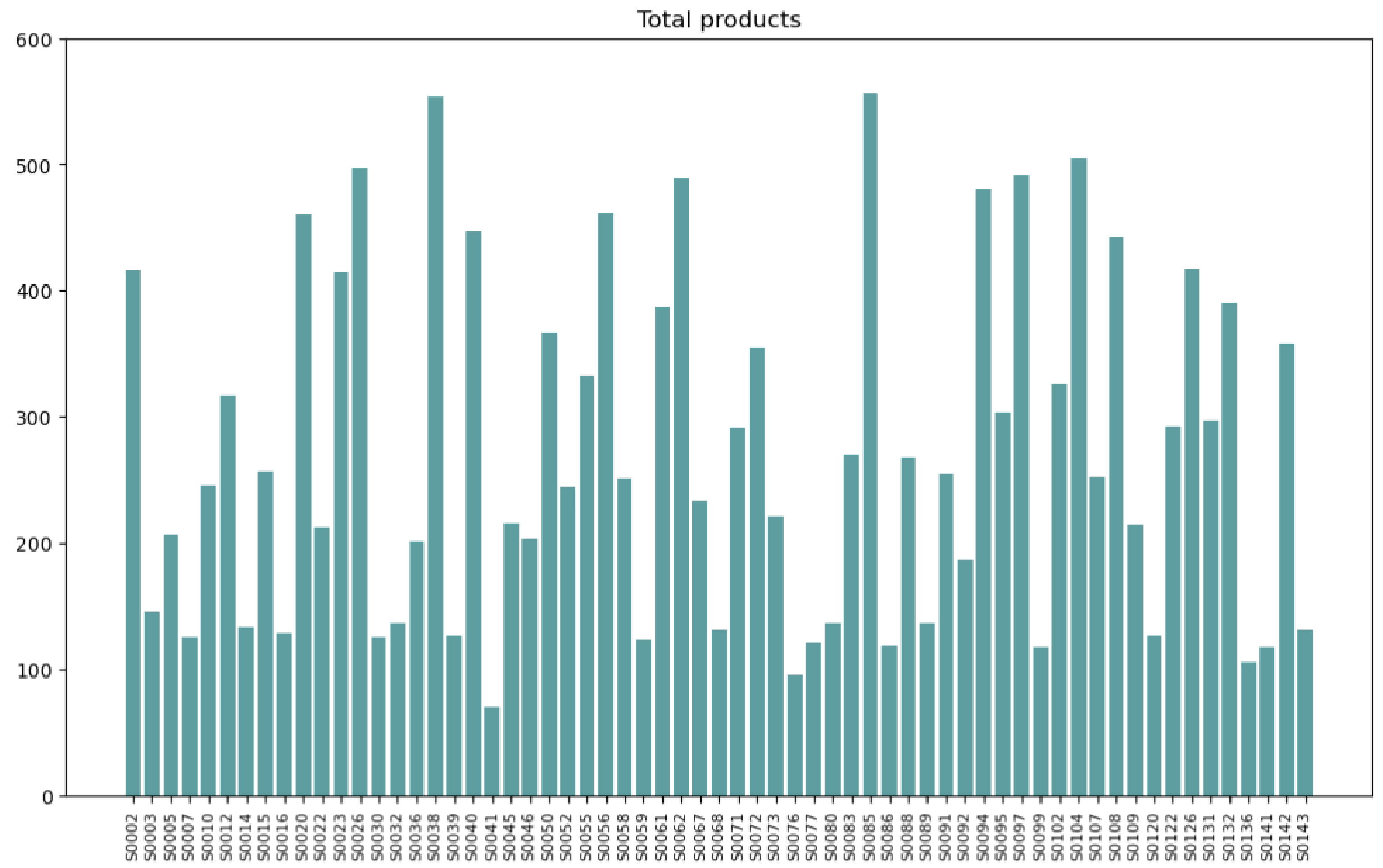
- Product ID
- Comprimento
- Largura
- Espessura

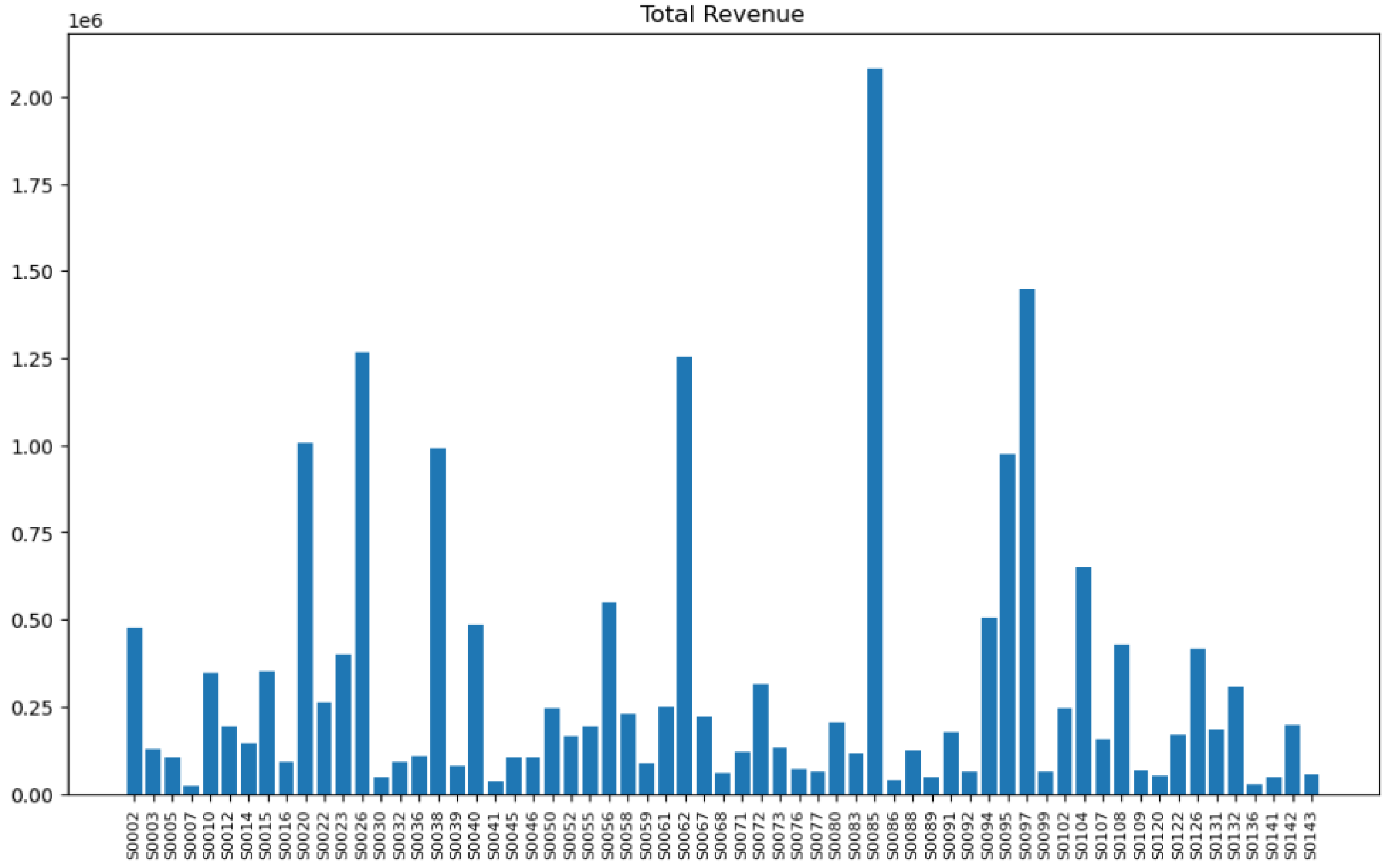
CITIES

- Store ID
- Tipo de loja
- Tamanho da loja
- País
- Cidade

63 lojas localizadas na Turquia

Dados de vendas de Janeiro 2017 a Setembro 2019





BUSINESS UNDERSTANDING

SALES
- Store ID
- Product ID
- Vendas
- Stock
- Preço
- Receitas
- Promoções

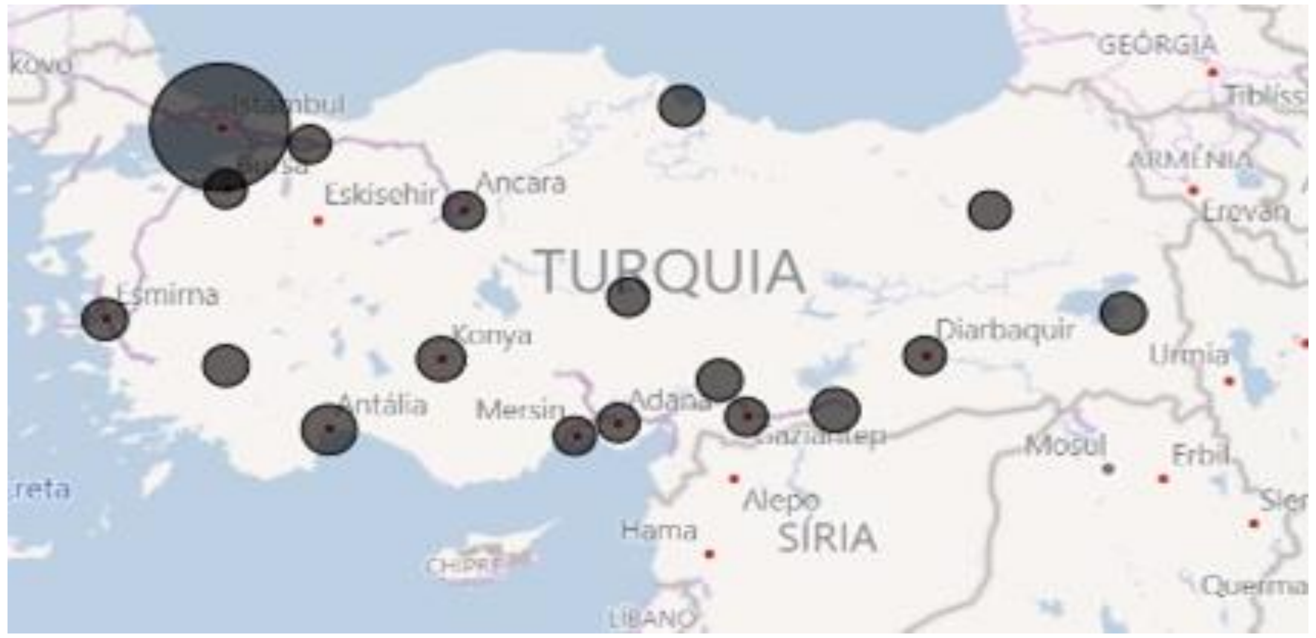
PRODUCT
- Product ID
- Comprimento
- Largura
- Espessura

CITIES
- Store ID
- Tipo de loja
- Tamanho da loja
- País
- Cidade

63 lojas localizadas na Turquia

Dados de vendas de Janeiro 2017 a Setembro 2019

Cidade	Lojas	População 2019
Istanbul	32	15.030.000
Antalya	5	2.420.000
Konya	3	2.160.000
Sanliurfa	3	2.073.000
Denizli	2	1.039.000
Izmir	2	3.760.000
Kahramanmaras	2	639.000
Samsun	2	1.340.000
Van	2	444.000
Adana	1	2.250.000
Adapazari	1	508.000
Ankara	1	5.500.000
Bursa	1	2.990.000
Diyarbakir	1	1.750.000
Erzurum	1	766.000
Eskiehir	1	875.000
Gaziantep	1	2.070.000
Kayseri	1	1.383.000
Mersin	1	1.750.000
Total	**63**	



Número de lojas em cada cidade

BUSINESS UNDERSTANDING

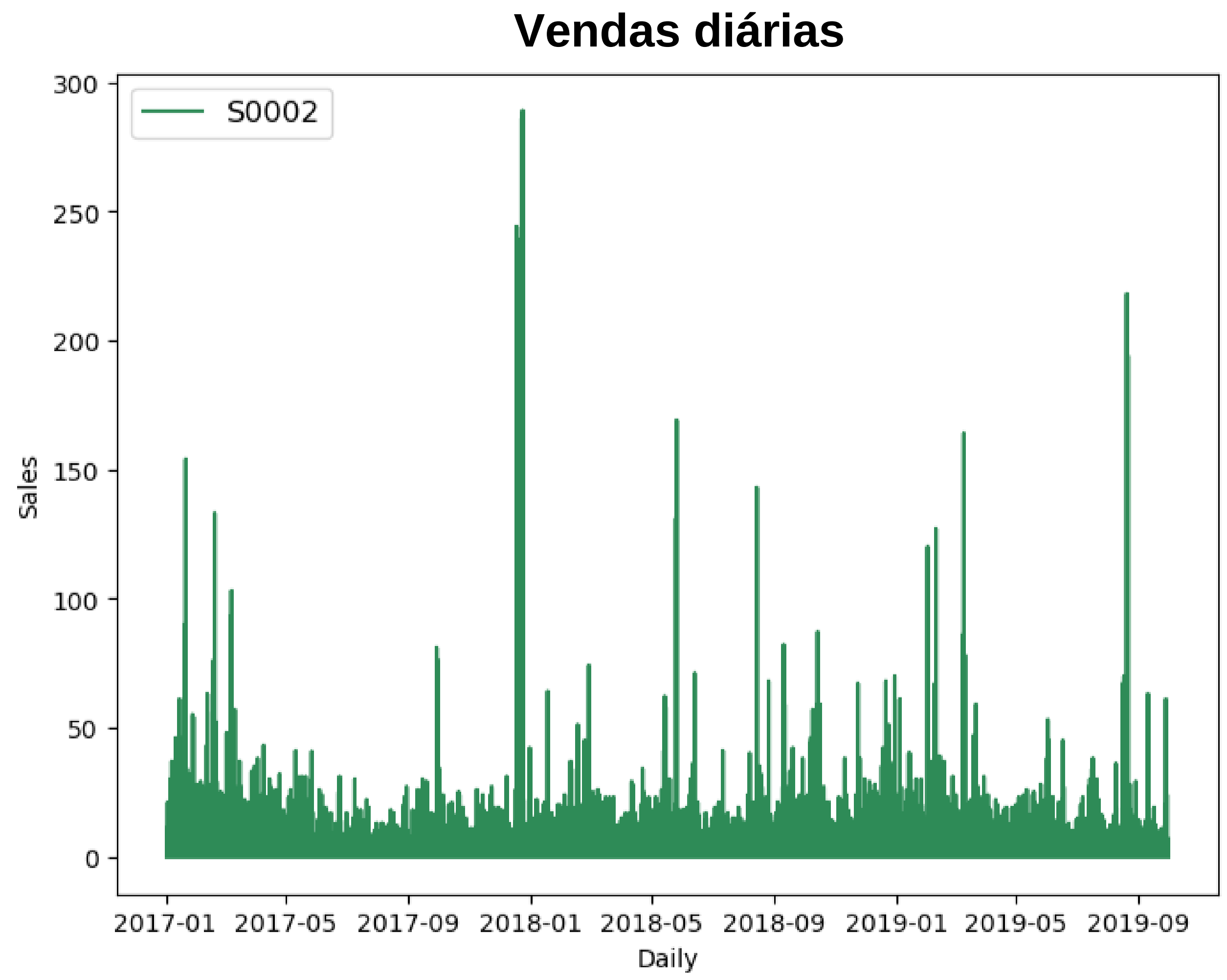
SALES
- Store ID
- Product ID
- Vendas
- Stock
- Preço
- Receitas
- Promoções

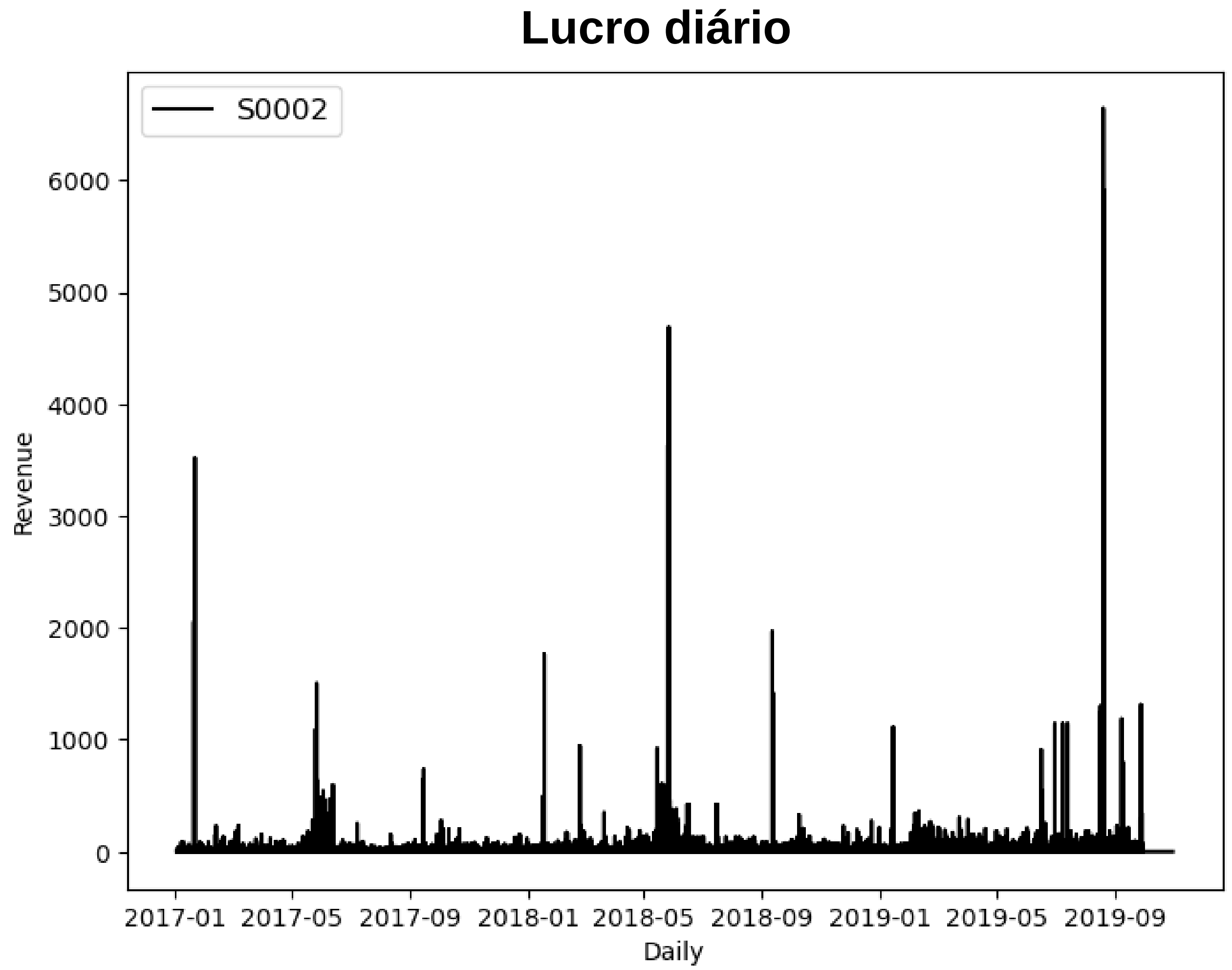
PRODUCT
- Product ID
- Comprimento
- Largura
- Espessura

CITIES
- Store ID
- Tipo de loja
- Tamanho da loja
- País
- Cidade

63 lojas localizadas na Turquia

Dados de vendas de Janeiro 2017 a Setembro 2019

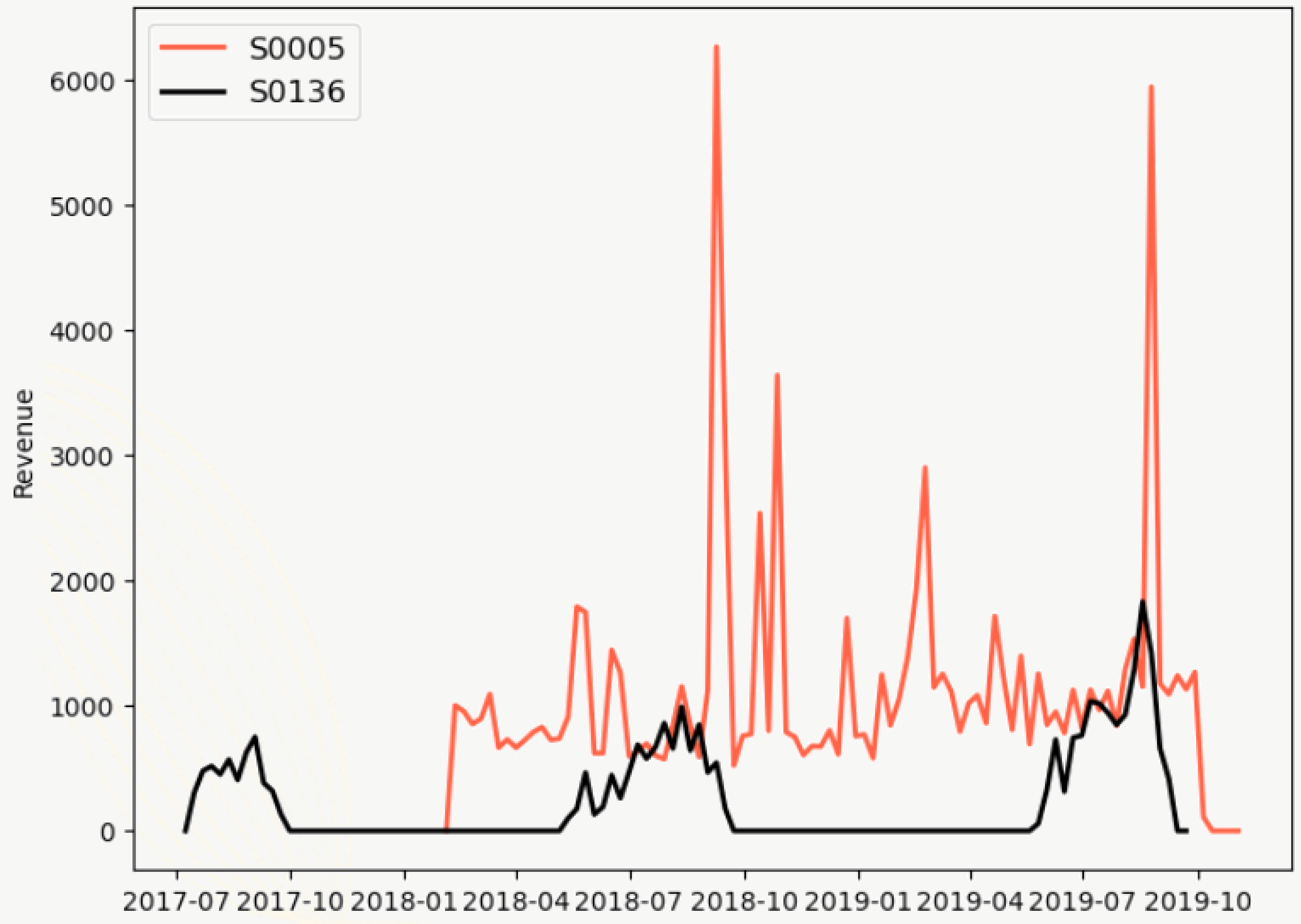




NOTA

O número de lojas ao longo do período sofreu flutuações:

- Abertura
- Sazonalidade



Objetivo: FORECAST

Previsão dos lucros de cada loja para o mês de Outubro 2019

DATA UNDERSTANDING / DATA PREPARATION

Data Profiling

Table	Sales
Número Variáveis	13
Linhas	8886058
Valores em falta	30.5%
Duplicados	0
Texto	2
Data	1
Variável Categórica	8
Variável Numérica	0

Vendas diárias

	store_id	product_id	date	sales	revenue	stock	price	promo_type_1	promo_bin_1	promo_type_2	promo_bin_2	promo_discount_2
1	S0002	P0001	2017-01-02	0.0	0.00	8.0	6.25	PR14	NaN	PR03	NaN	NaN
2	S0002	P0005	2017-01-02	0.0	0.00	11.0	33.90	PR14	NaN	PR03	NaN	NaN
3	S0002	P0011	2017-01-02	0.0	0.00	9.0	49.90	PR14	NaN	PR03	NaN	NaN
4	S0002	P0015	2017-01-02	1.0	2.41	19.0	2.60	PR14	NaN	PR03	NaN	NaN
5	S0002	P0017	2017-01-02	0.0	0.00	12.0	1.49	PR14	NaN	PR03	NaN	NaN

Total rows: 8886058

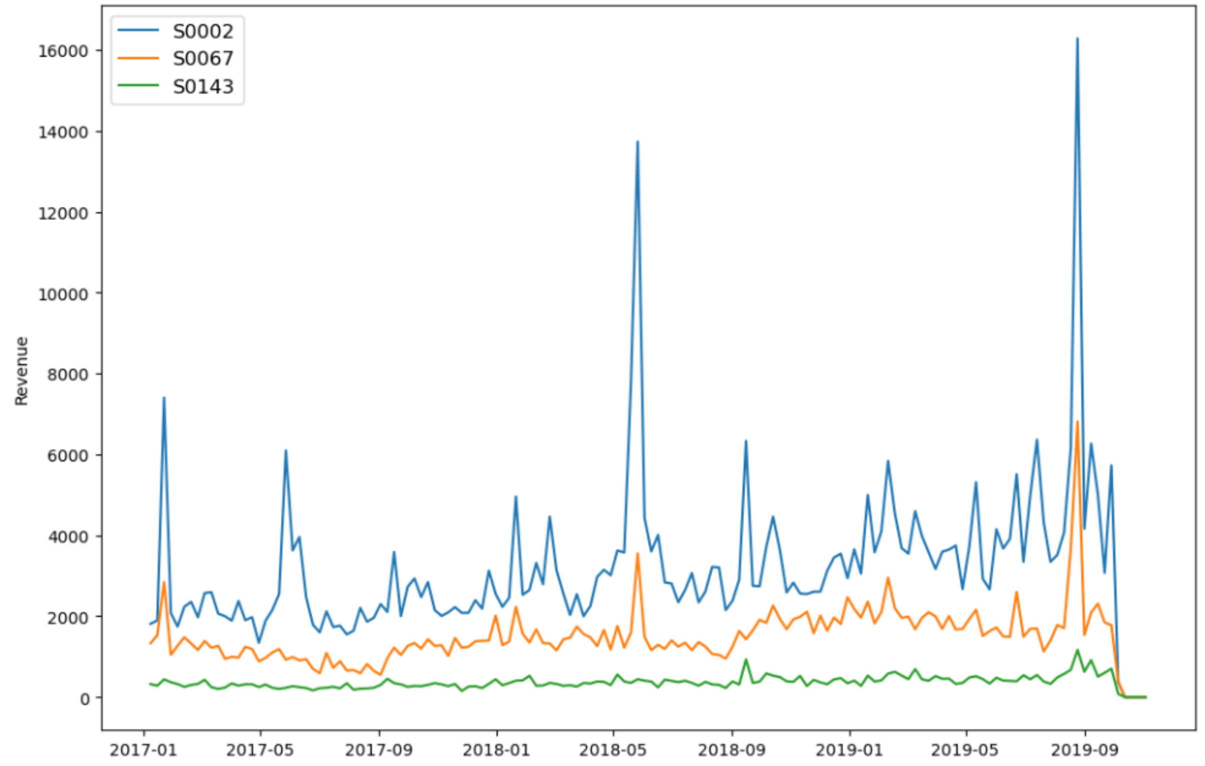
Agrupamento semanal das vendas

vendas semanais

date	store_id	sales	revenue
2017-01-08	S0002	750.232	1813.24
2017-01-15	S0002	871.530	1896.91
2017-01-22	S0002	1069.555	7404.16
2017-01-29	S0002	910.255	2078.39
2017-02-05	S0002	732.754	1745.47
...	...	...	...
2019-10-06	S0143	18.000	80.82
2019-10-13	S0143	0.000	0.00
2019-10-20	S0143	0.000	0.00
2019-10-27	S0143	0.000	0.00
2019-11-03	S0143	0.000	0.00

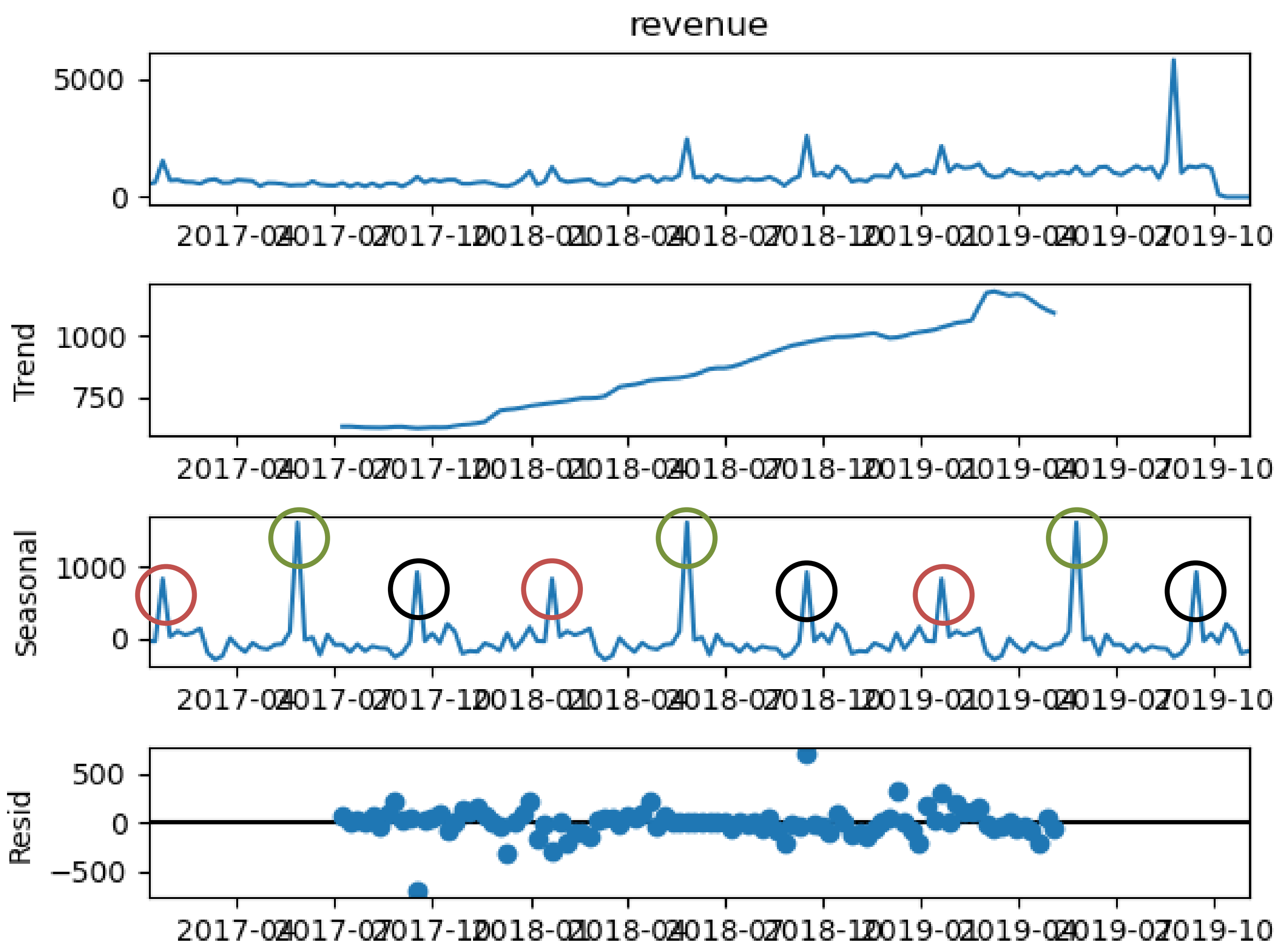
Forecast

Total rows: 8610

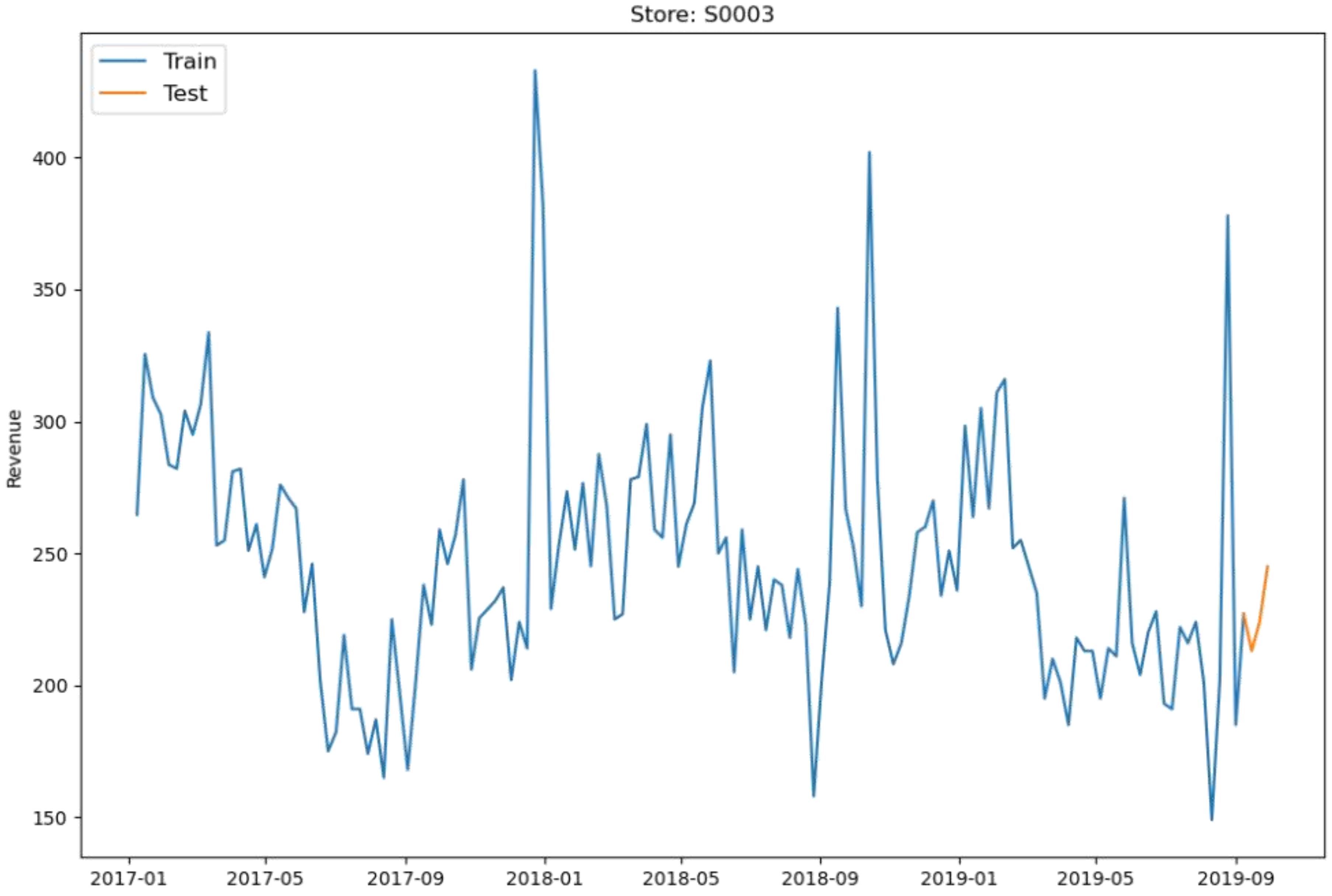


DATA UNDERSTANDING / DATA PREPARATION

Decomposição da série temporal

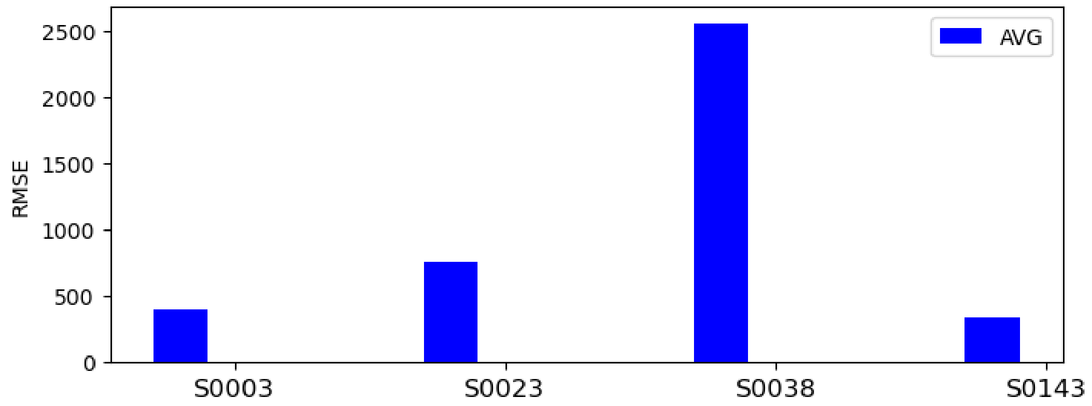


Train vs Test

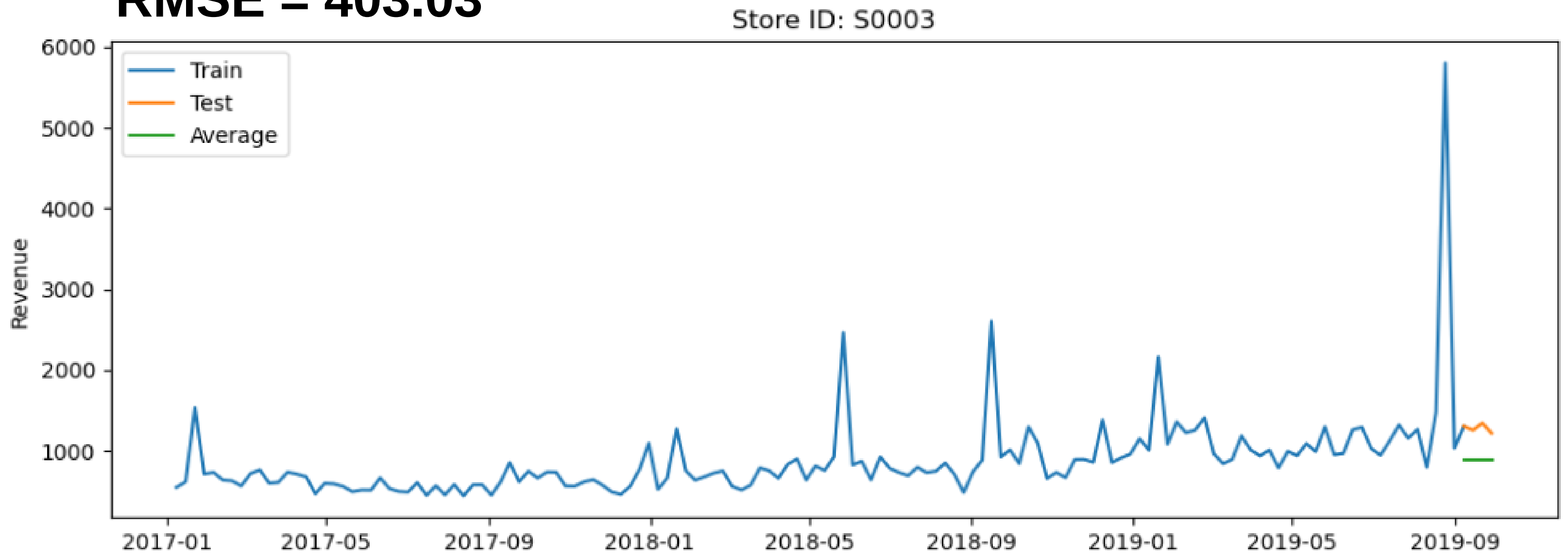


Treino	2017-01-08 : 2019-09-08	**140 semanas**
Teste	2019-09-08 : 2019-09-29	**4 semanas**

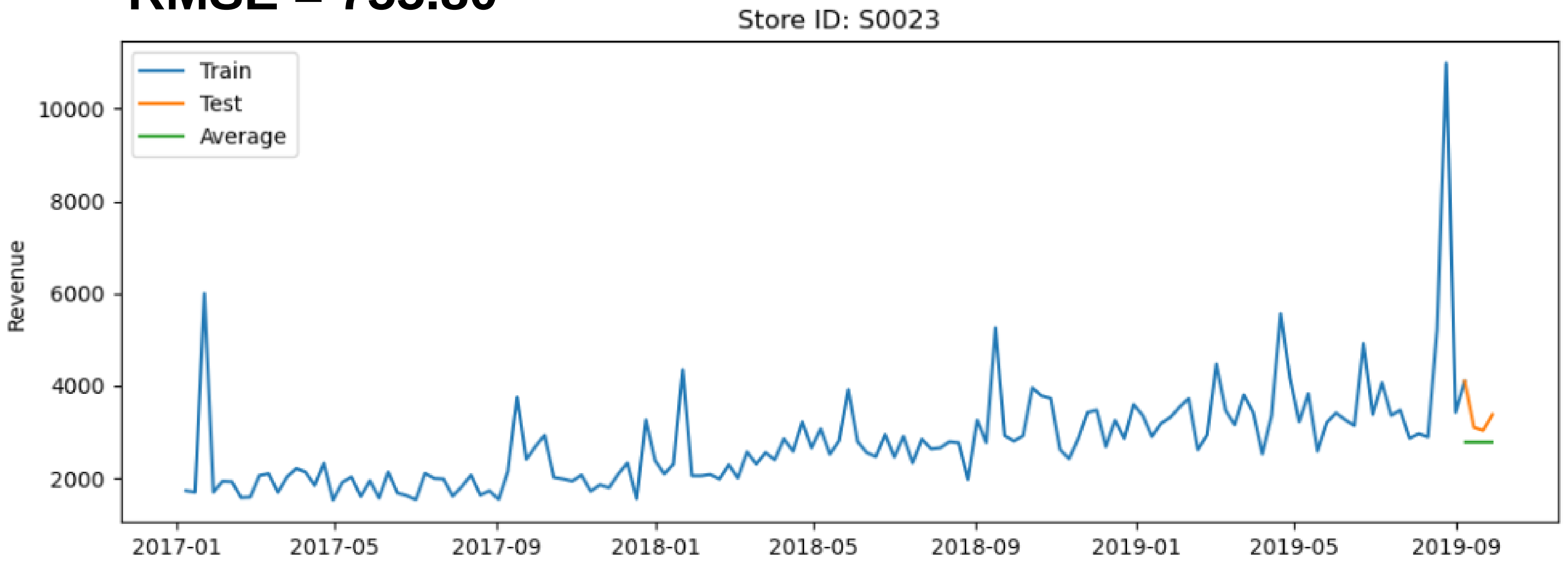
MODELING – SIMPLE AVERAGE



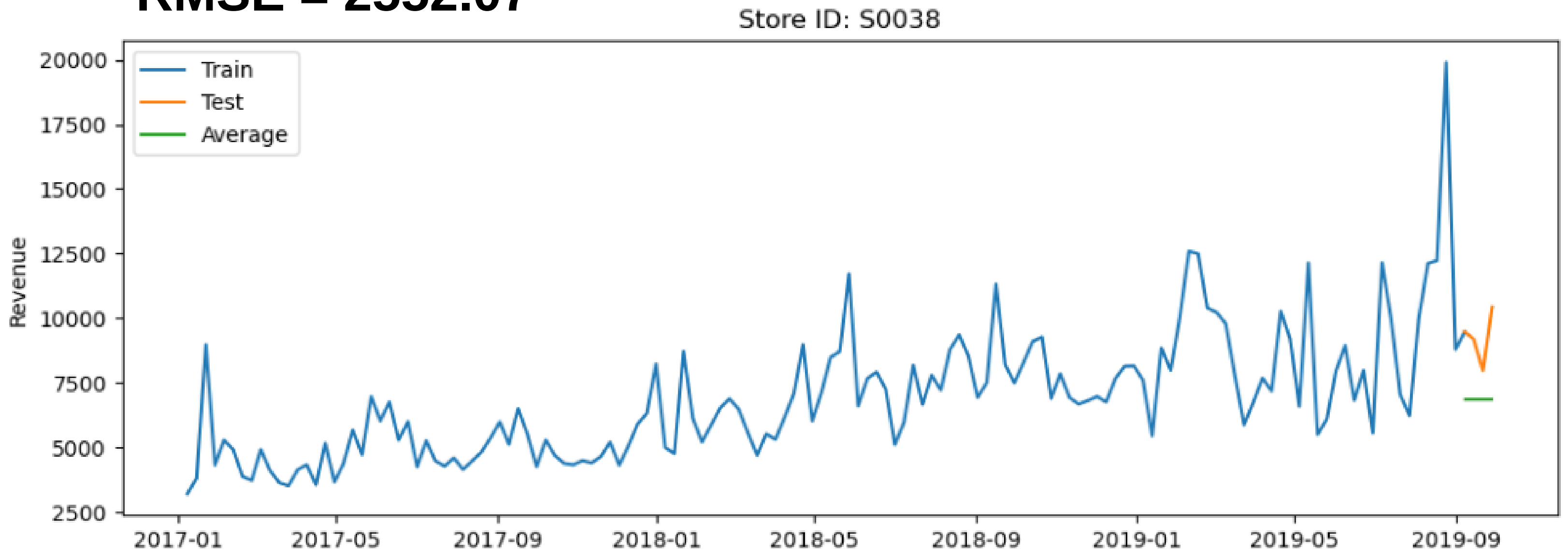
RMSE = 403.03



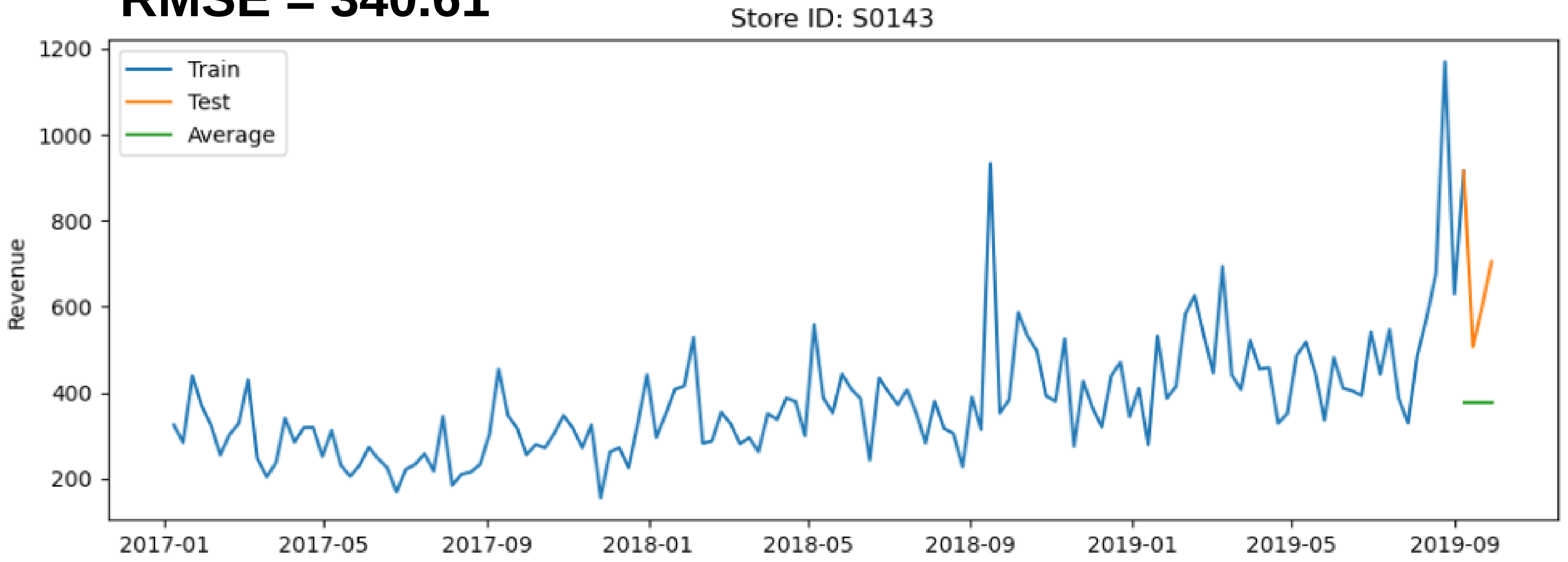
RMSE = 755.80



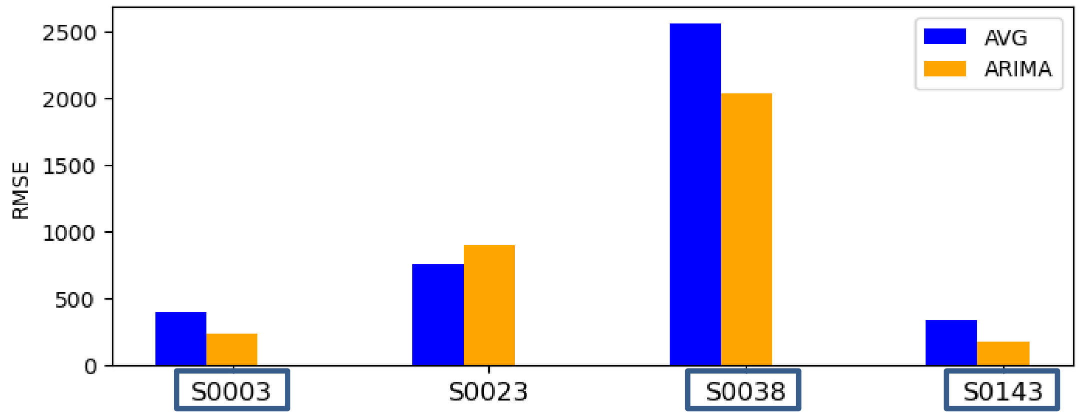
RMSE = 2552.07



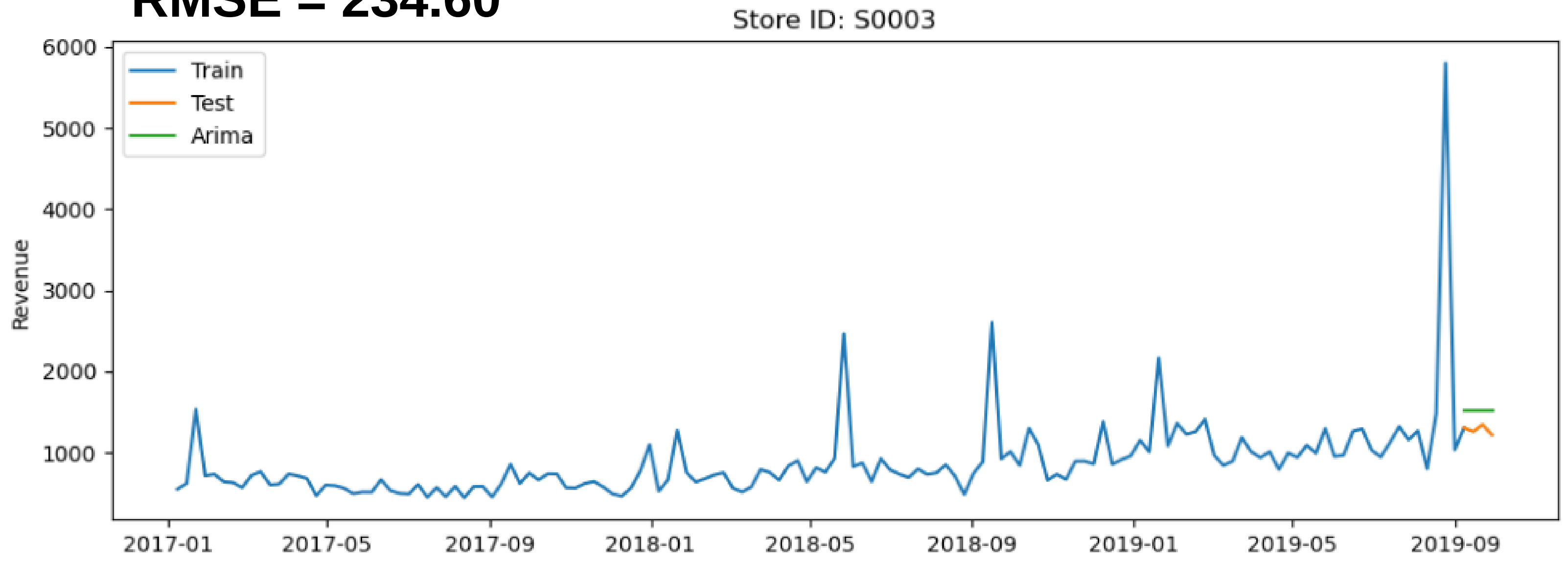
RMSE = 340.61



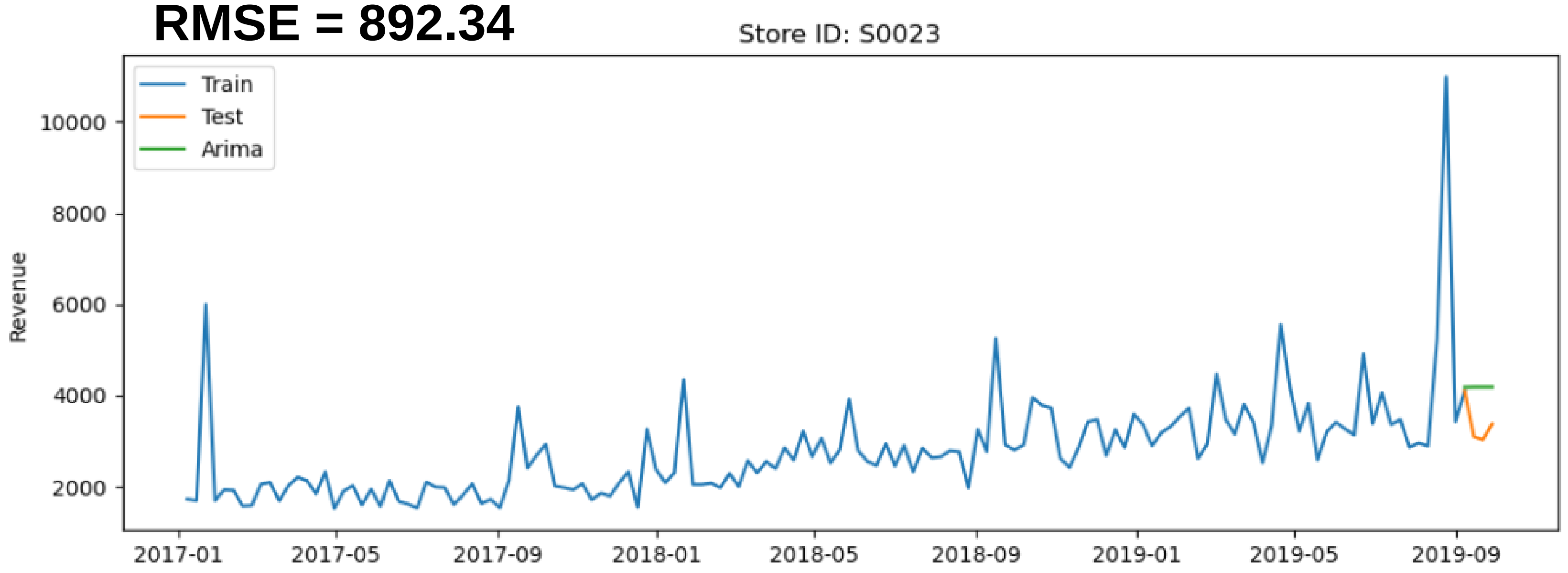
MODELING - ARIMA



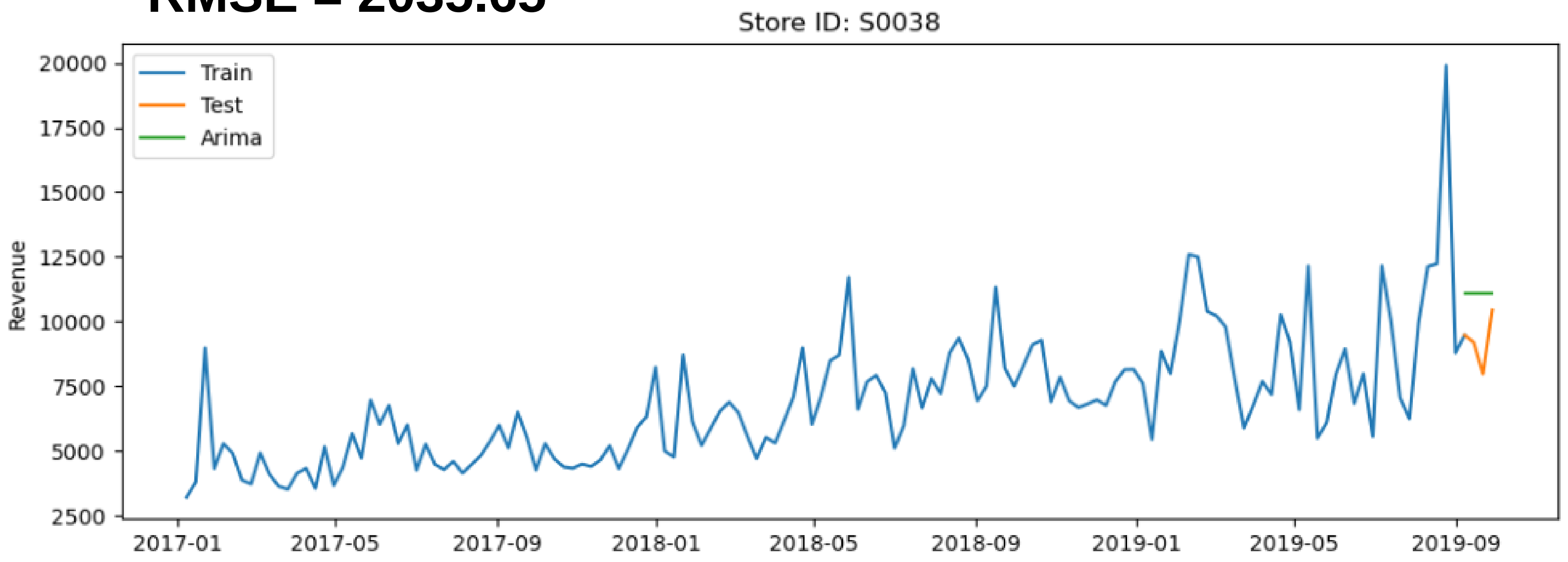
RMSE = 234.60



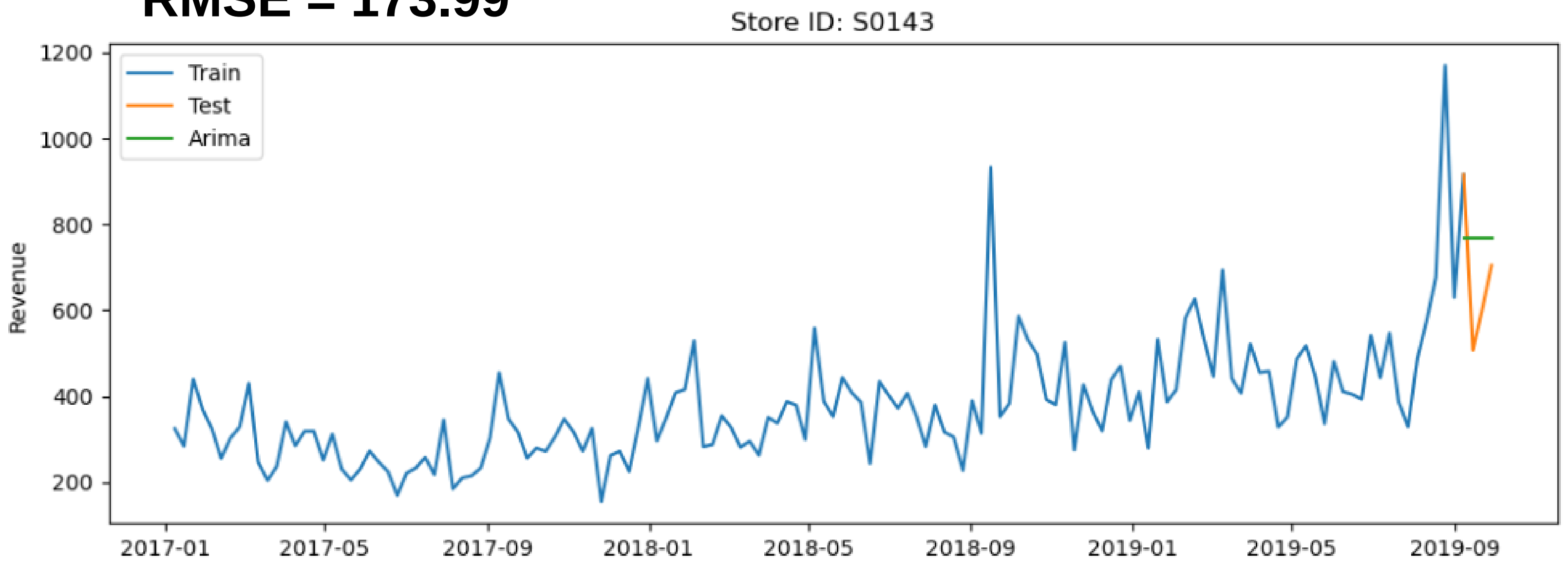
RMSE = 892.34



RMSE = 2035.65



RMSE = 173.99

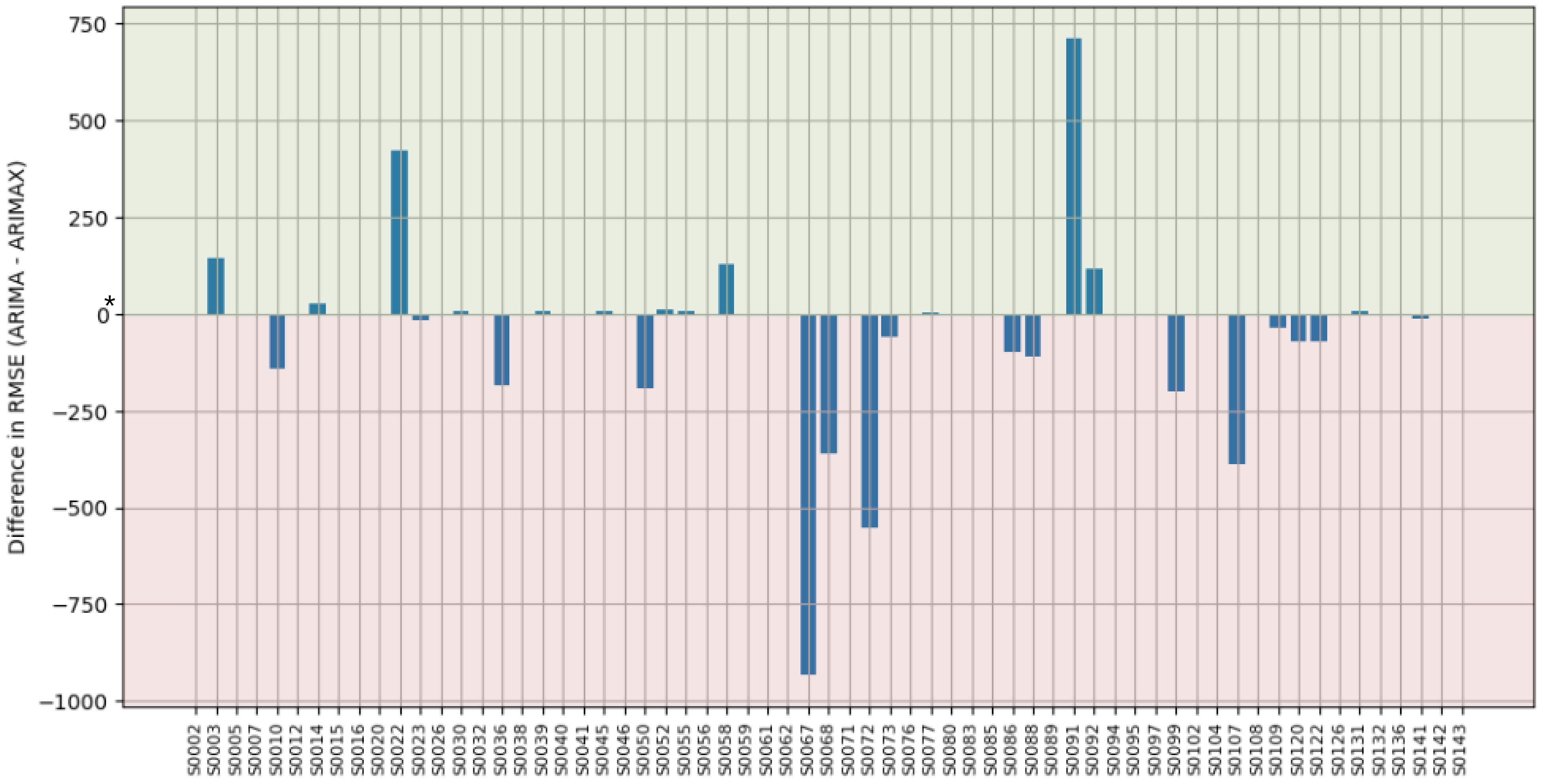


MODELING - ARIMAX

$\overbrace{\text{promo\_discount\_2} \quad \text{is\_holiday} \quad \text{month}}^{\text{Variáveis exógenas}}$

	store_id	sales	revenue	promo_discount_2	is_holiday	month
date						
2019-09-08	S0002	598.00	6267.91	1.0	0	9
2019-09-15	S0002	656.00	5007.57	1.0	0	9
2019-09-22	S0002	525.00	3074.56	1.0	0	9
2019-09-29	S0002	656.16	5728.25	1.0	0	9
2019-09-08	S0003	227.00	1302.61	1.0	0	9
...	...	...	...	...	...	...
2019-09-29	S0142	302.00	1866.41	1.0	0	9
2019-09-08	S0143	104.00	915.27	1.0	0	9
2019-09-15	S0143	63.00	507.08	1.0	0	9
2019-09-22	S0143	89.00	601.63	1.0	0	9
2019-09-29	S0143	147.00	704.87	1.0	0	9

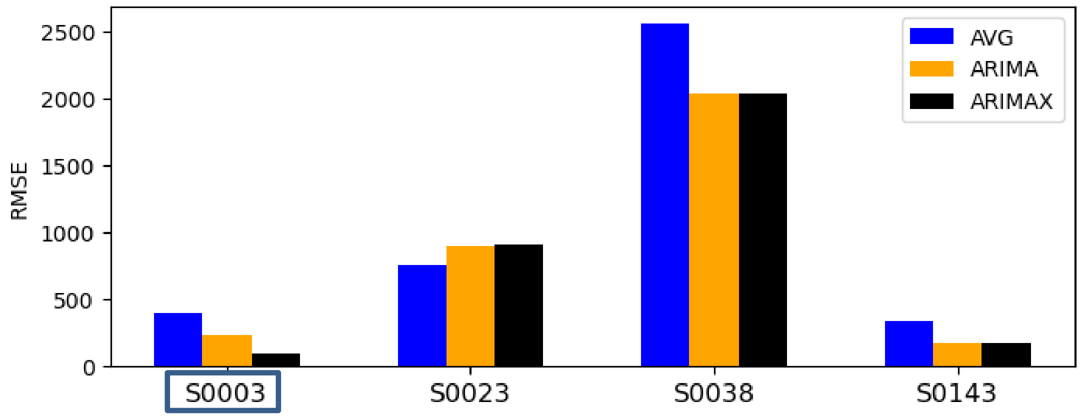
$RMSE_{arimax}$ < $RMSE_{arima}$ ➜ Efeito positivo das variáveis exógenas



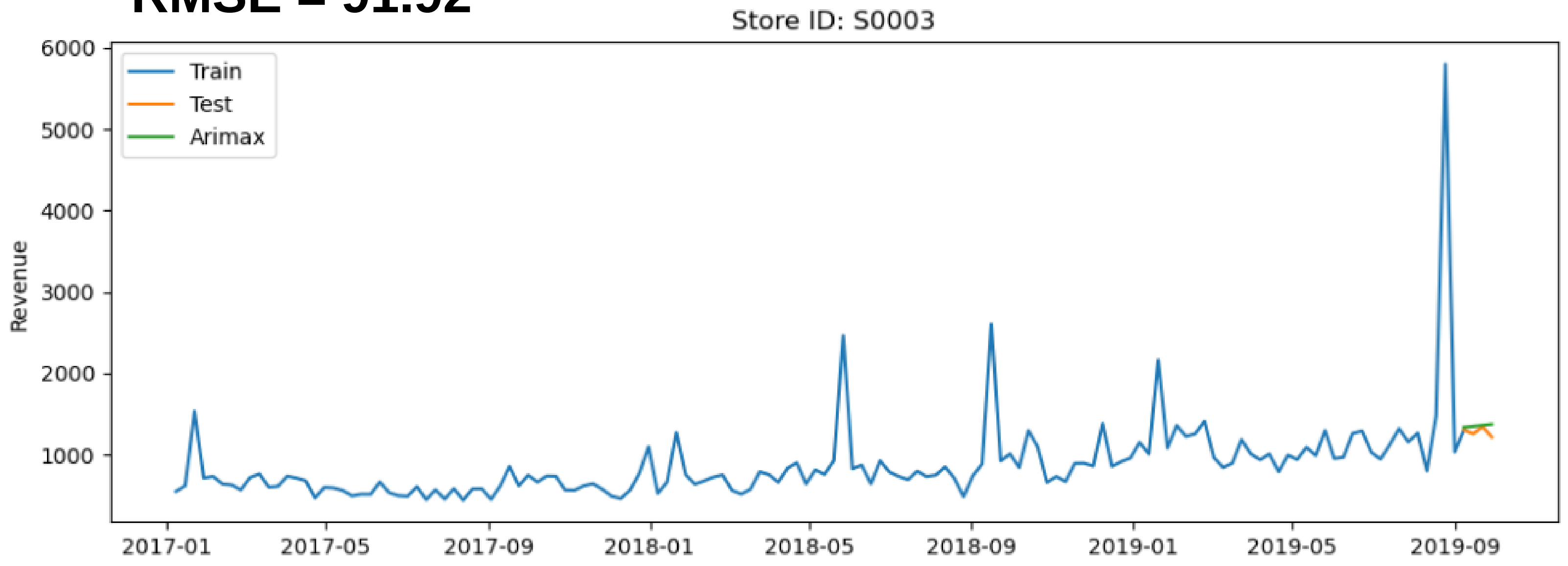
$RMSE_{arimax}$ > $RMSE_{arima}$ ➜ Efeito negativo das variáveis exógenas

* $RMSE_{arimax}$ = $RMSE_{arima}$ ➜ Variáveis exógenas não têm impacto

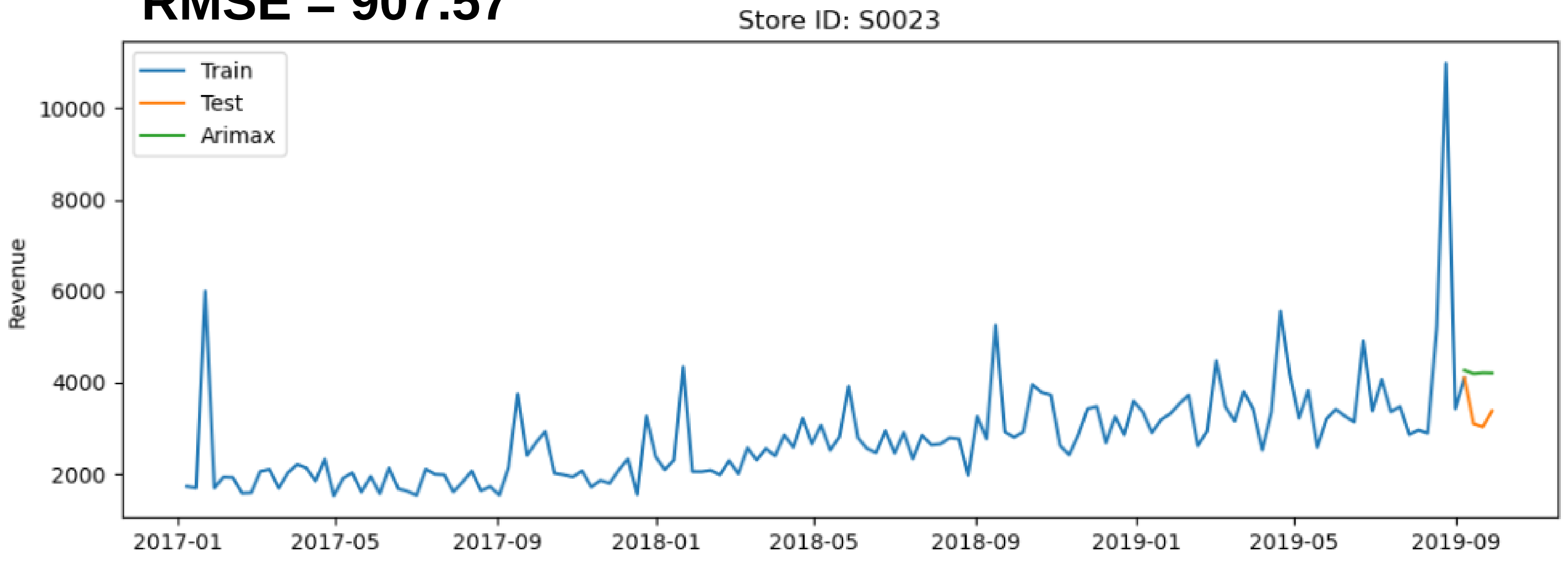
MODELING - ARIMAX



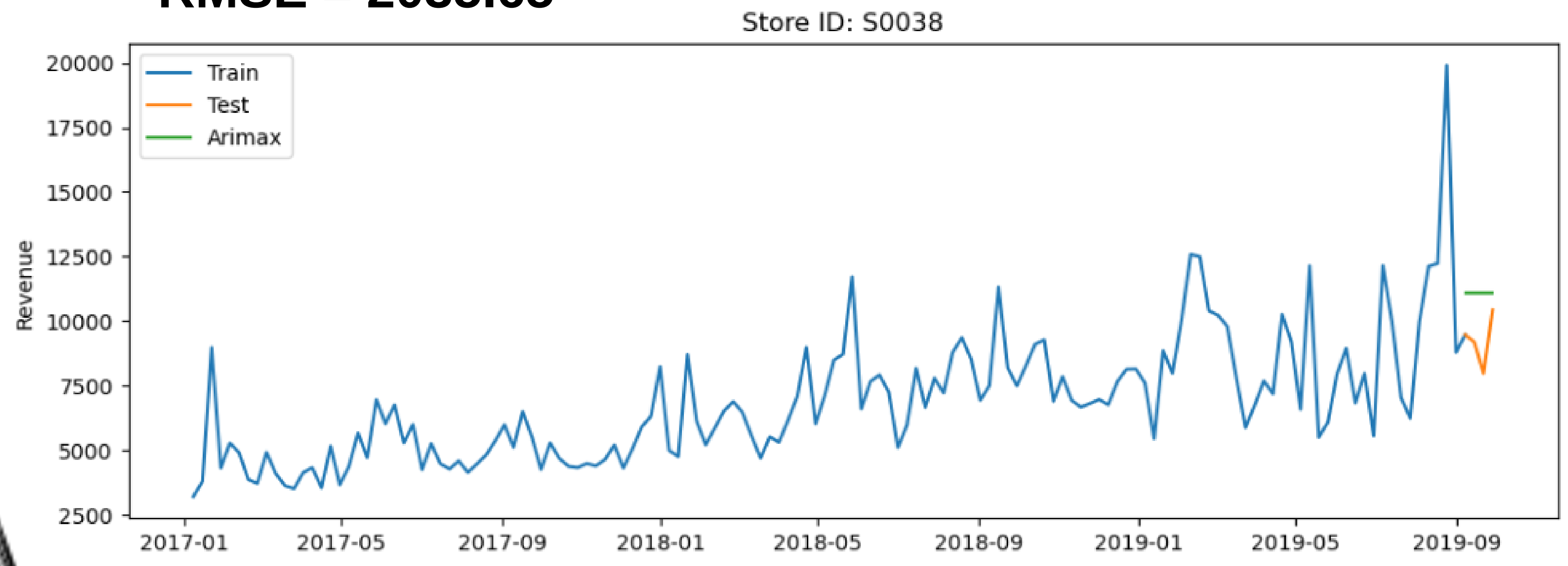
RMSE = 91.92



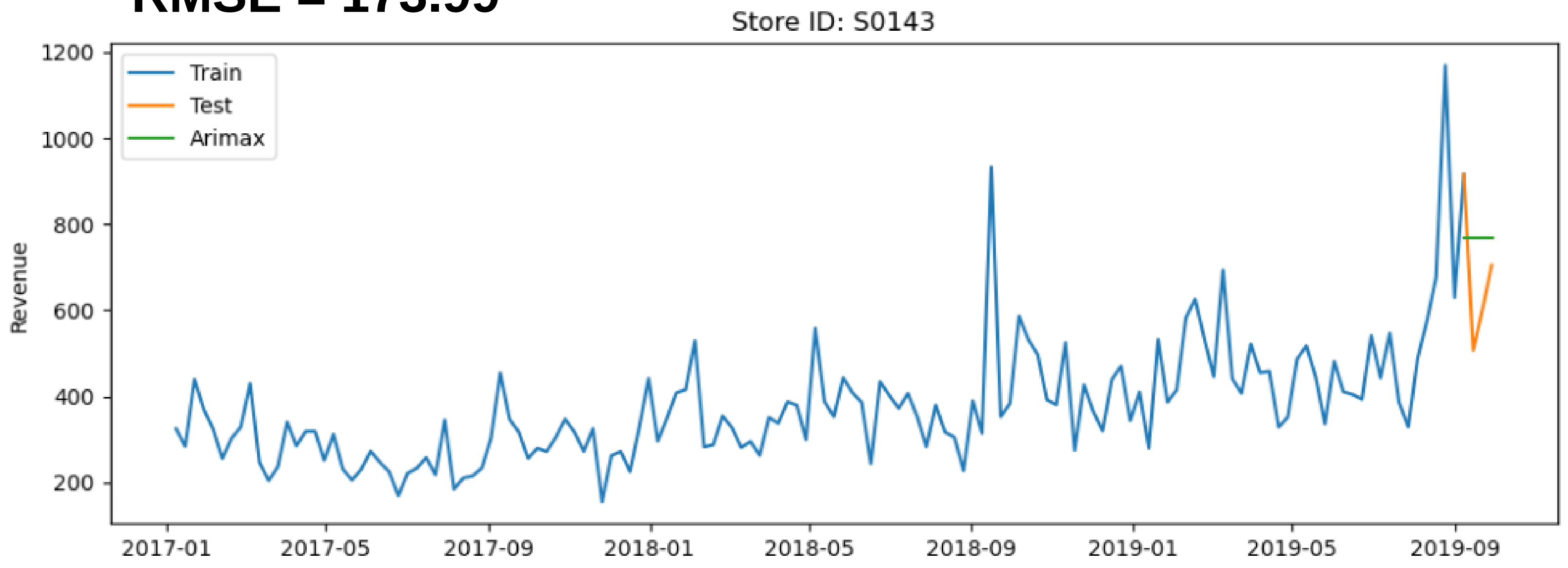
RMSE = 907.57



RMSE = 2035.65

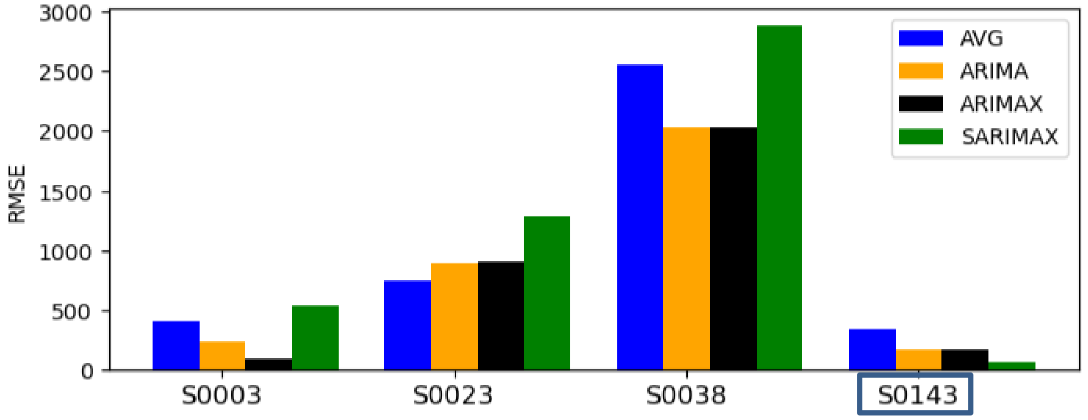


RMSE = 173.99

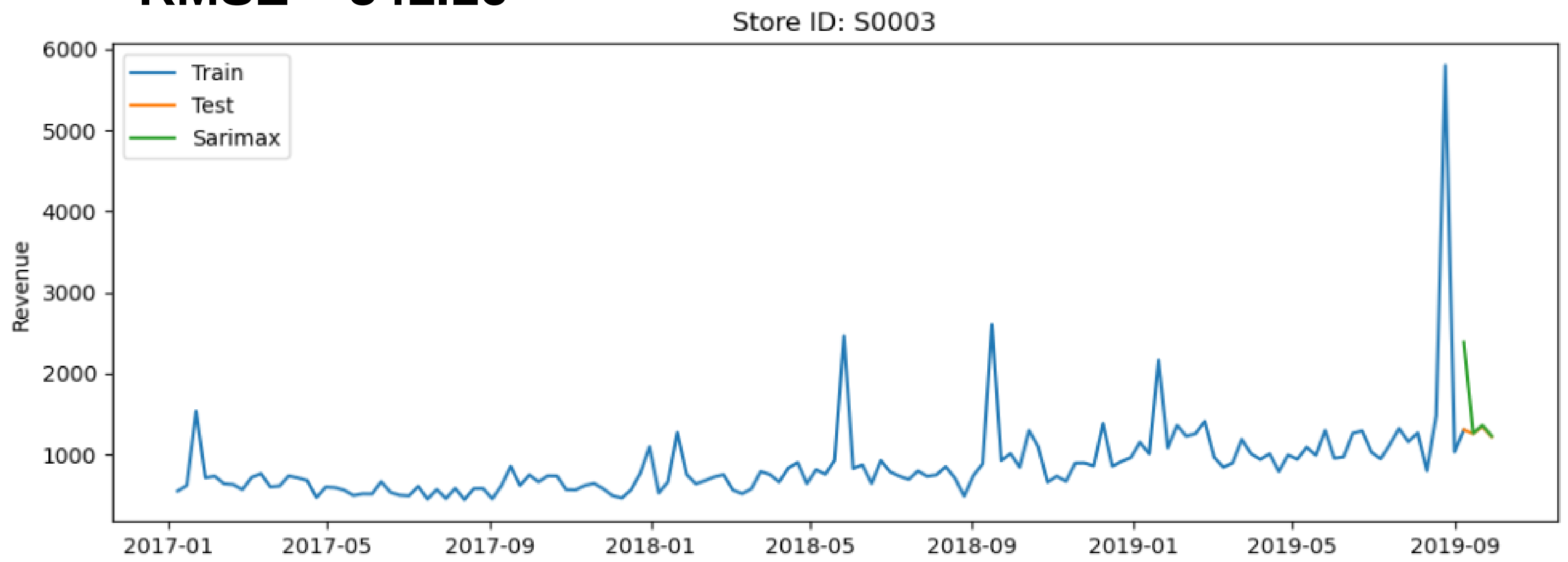


MODELING - SARIMAX

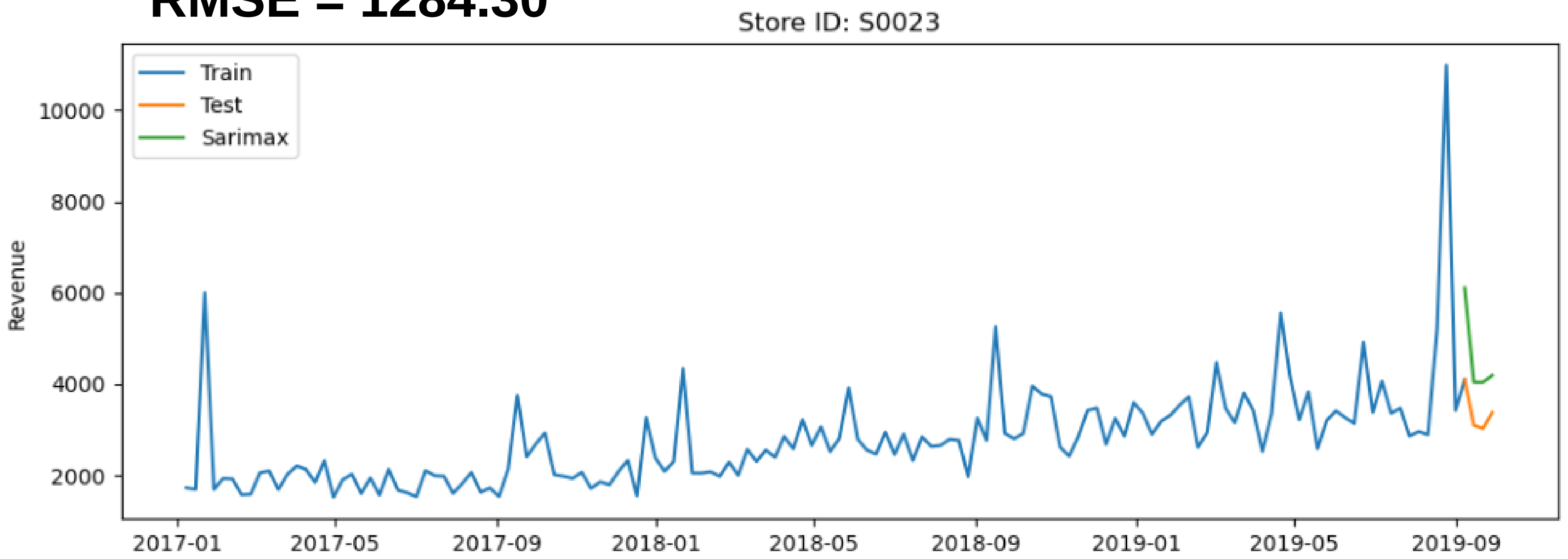
➢ **Introdução de sazonalidade no modelo**



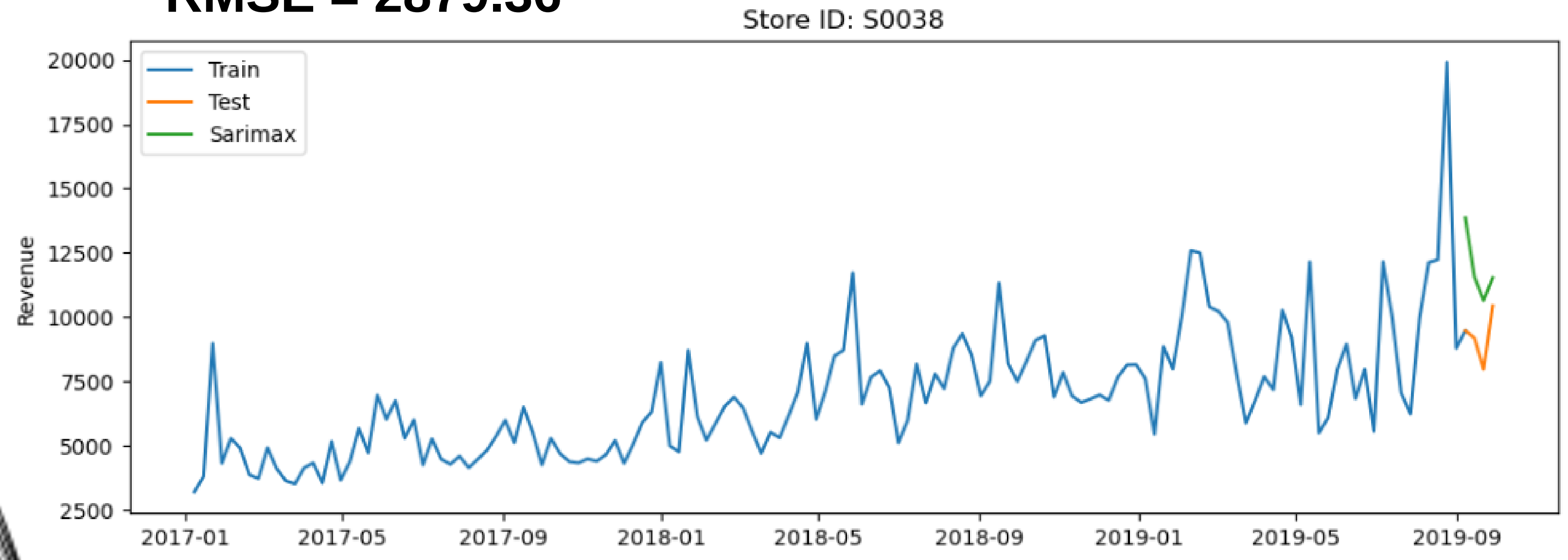
RMSE = 541.16



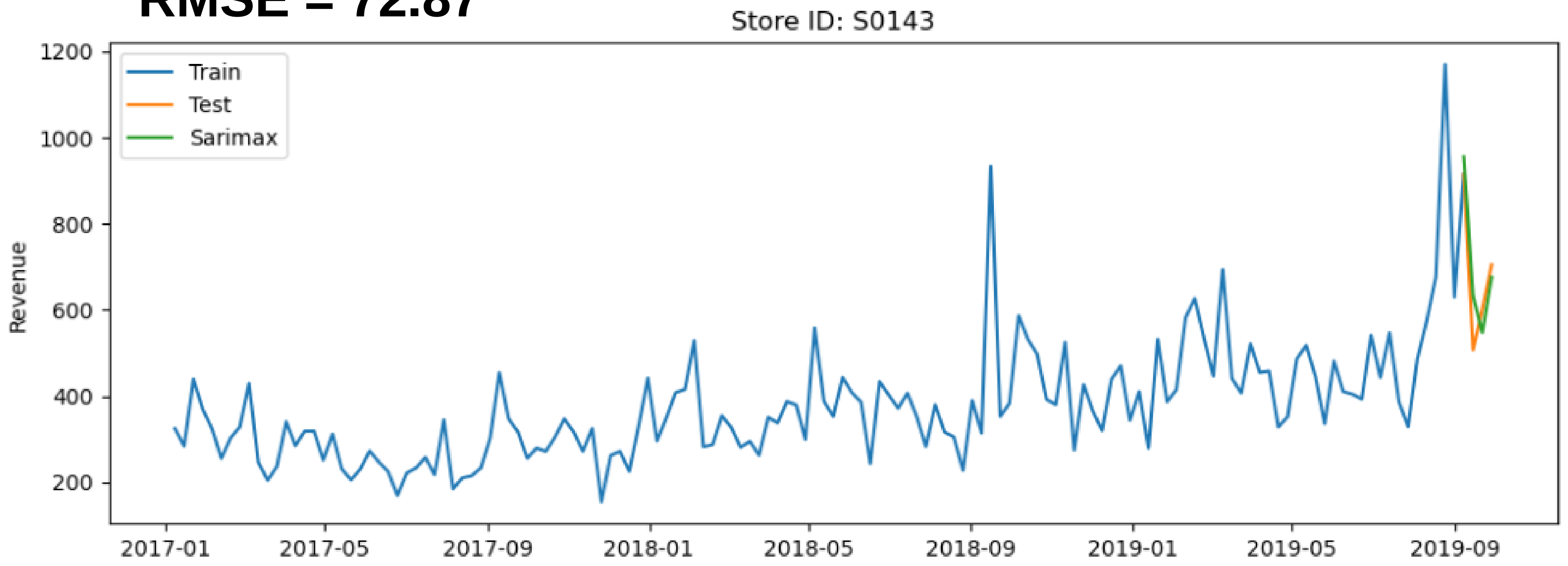
RMSE = 1284.30



RMSE = 2879.36

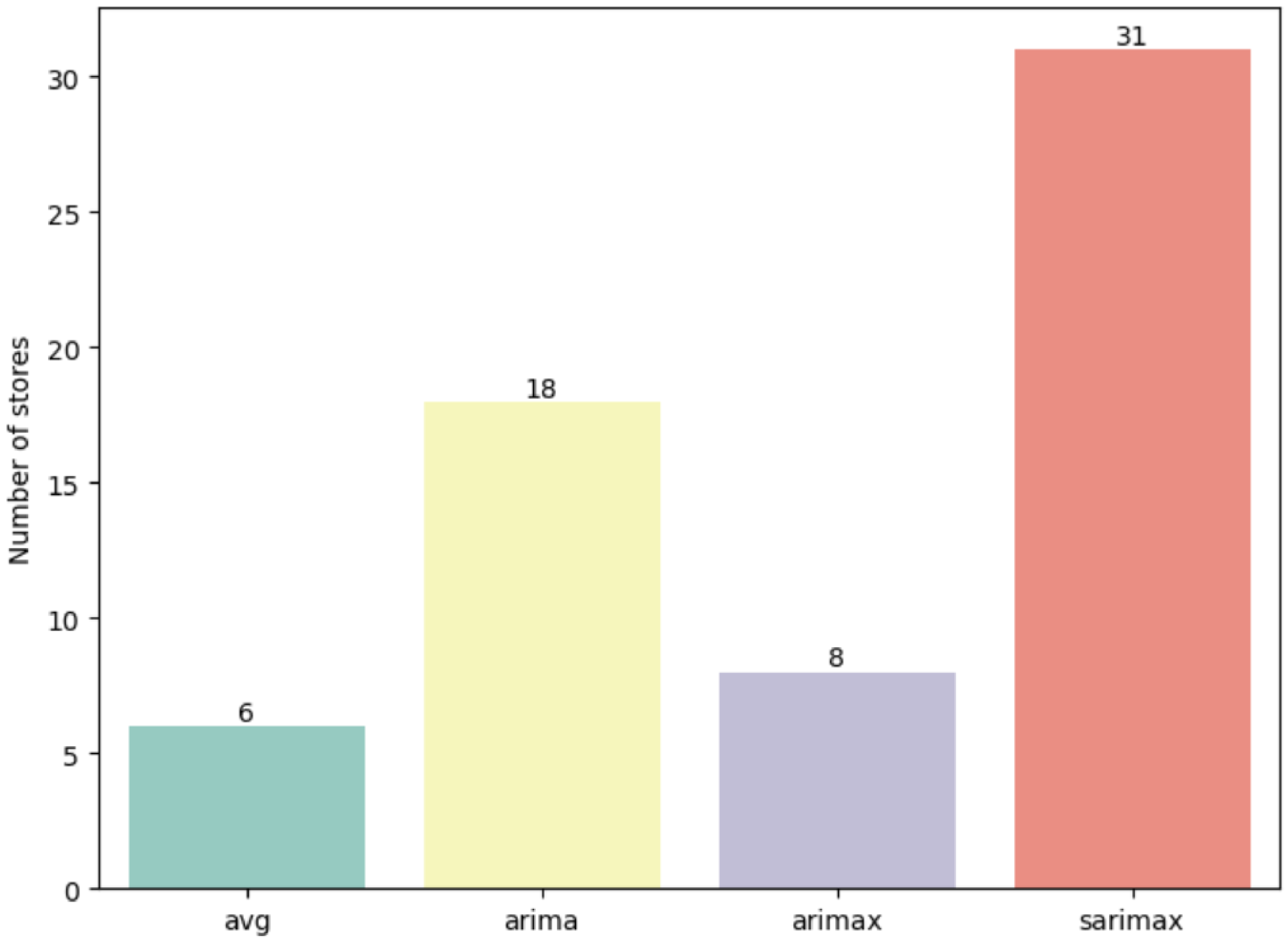


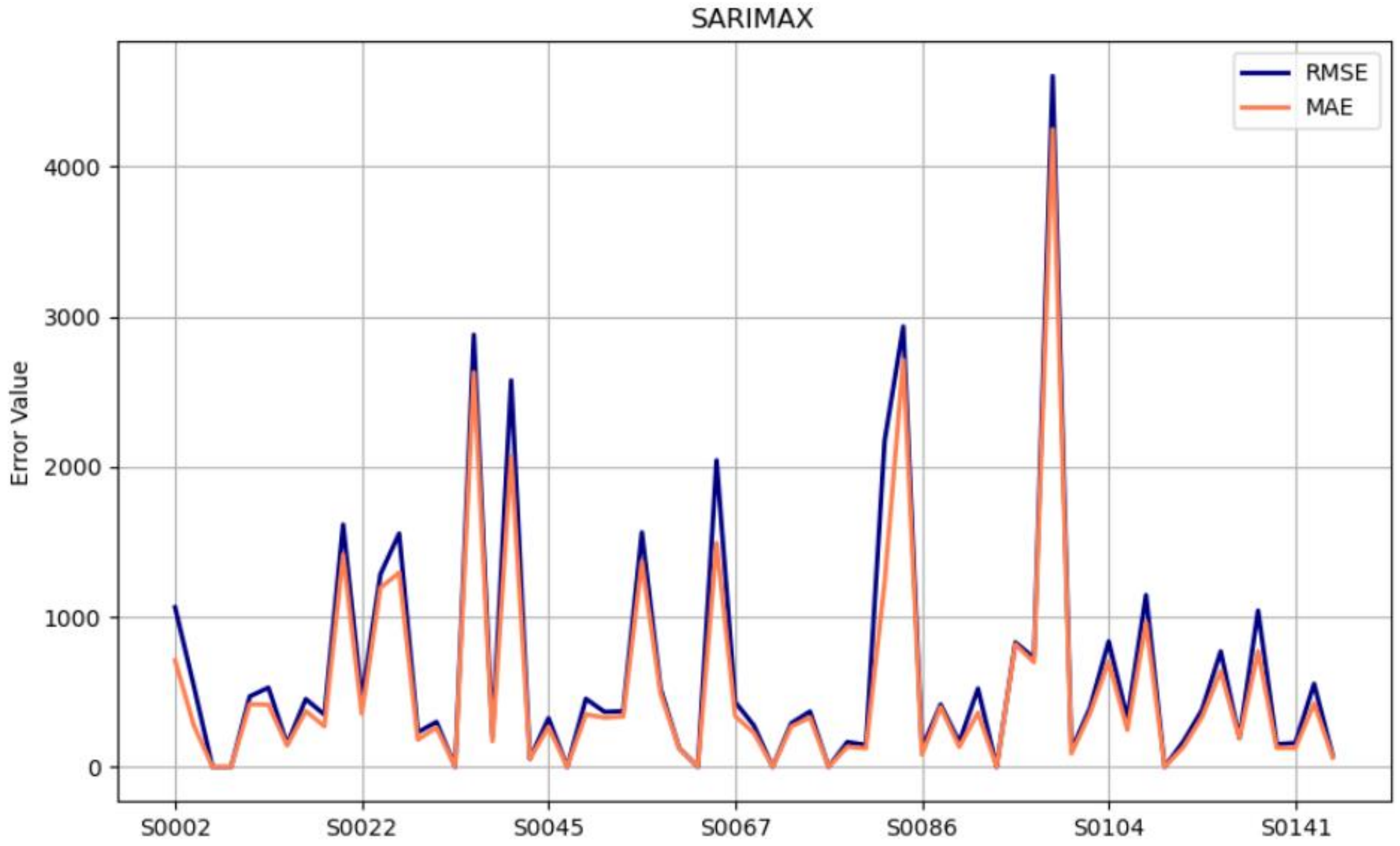
RMSE = 72.87

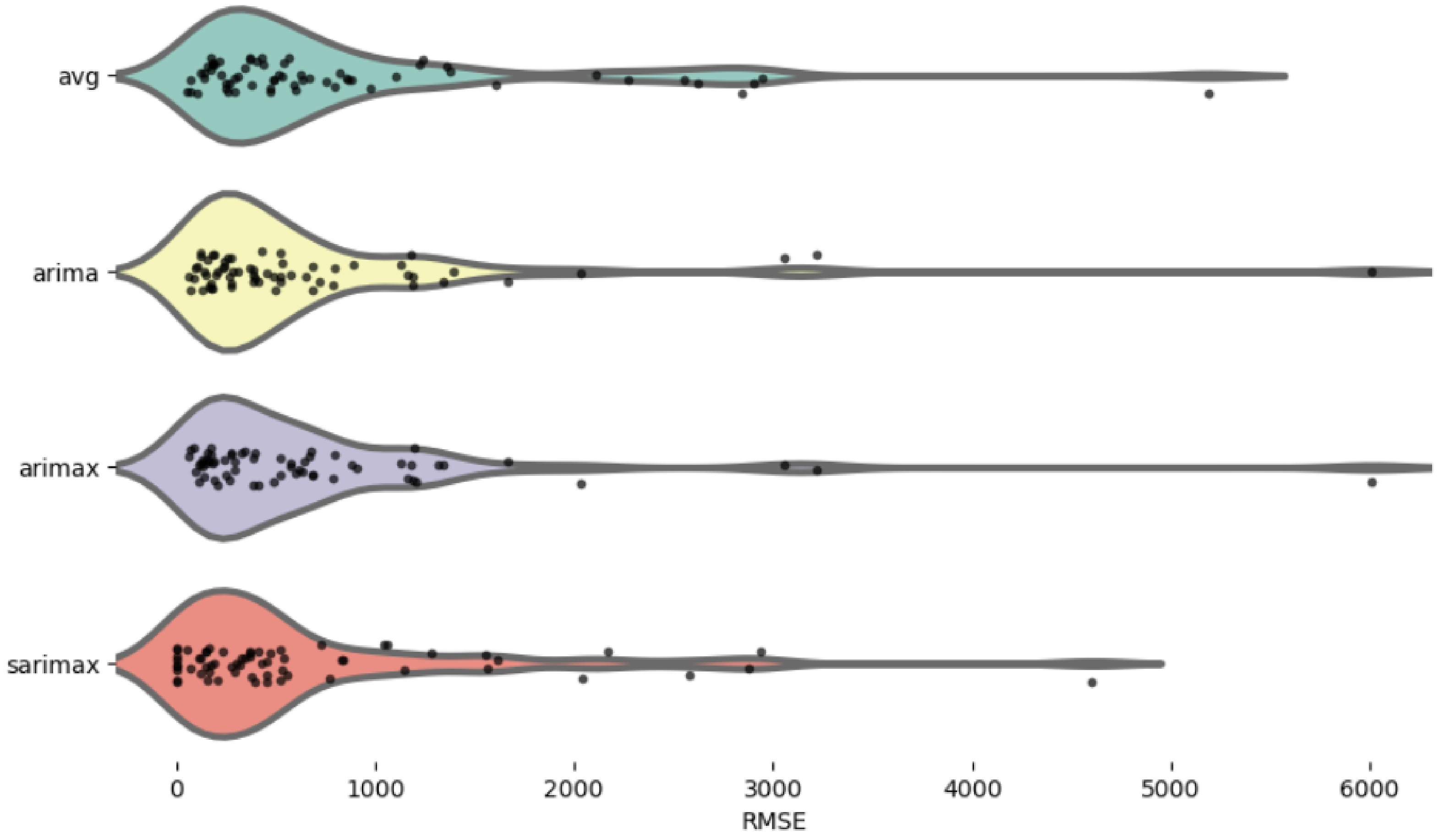


EVALUATION – MÉTRICAS

- Escolha dos melhores modelos de previsão para cada loja

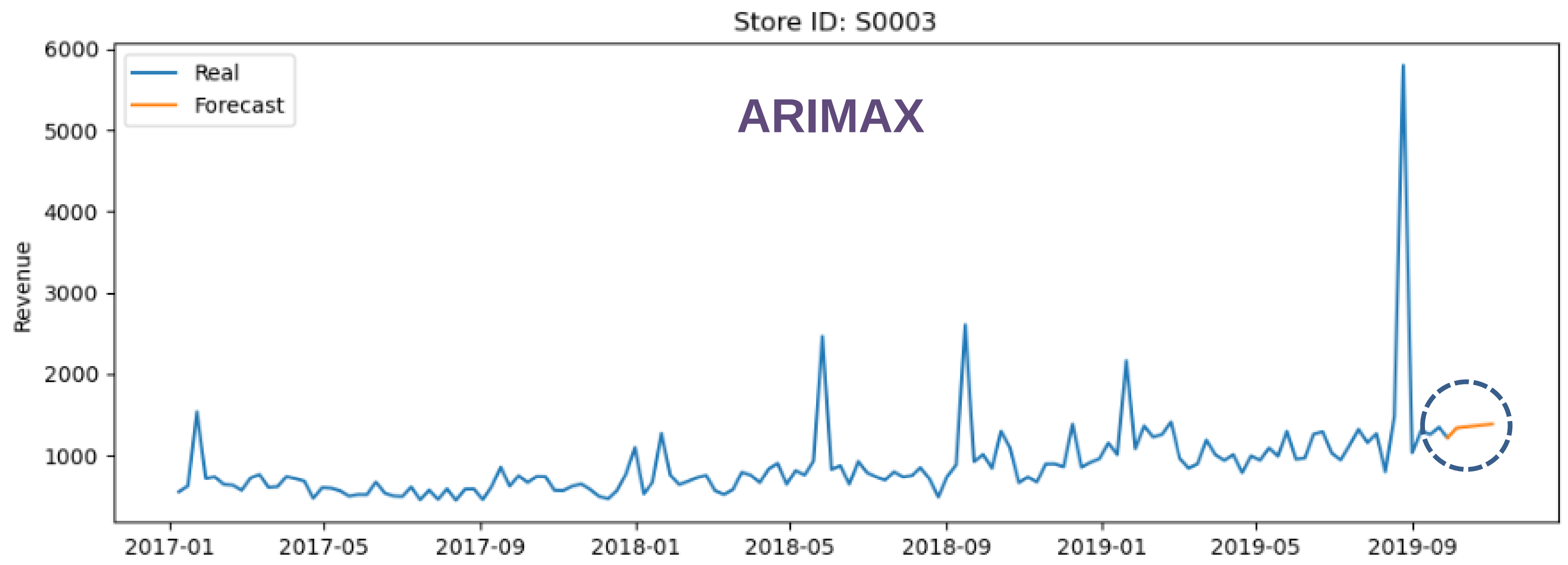


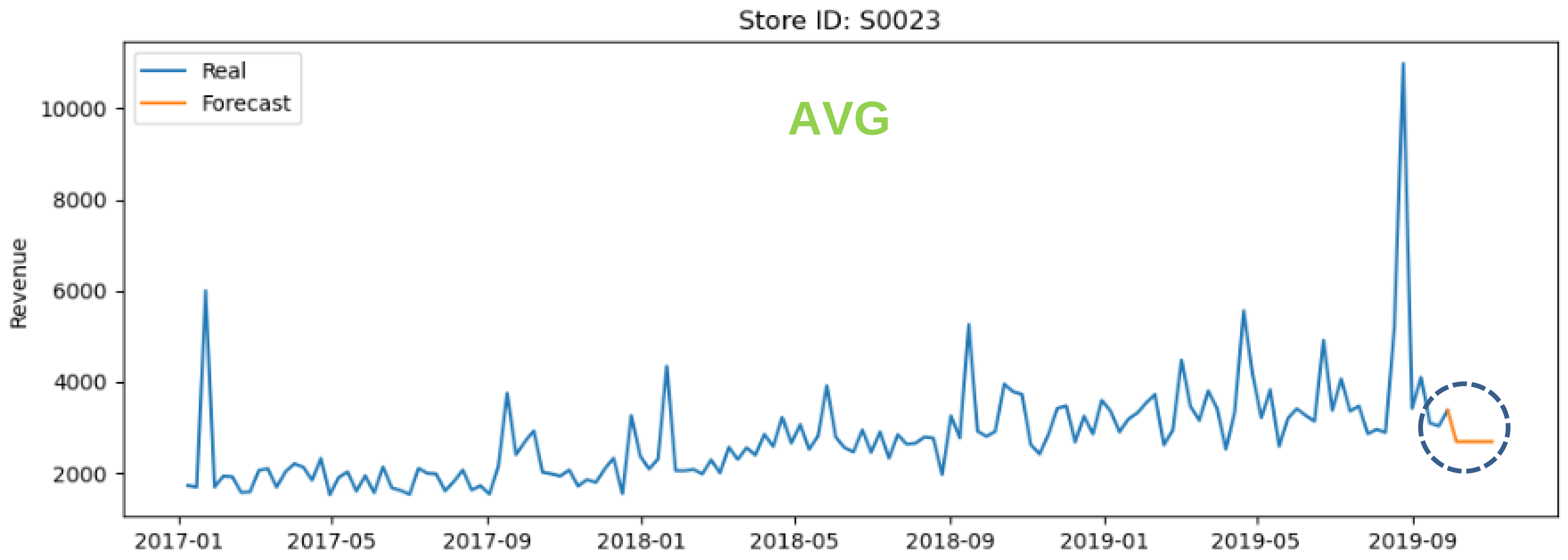


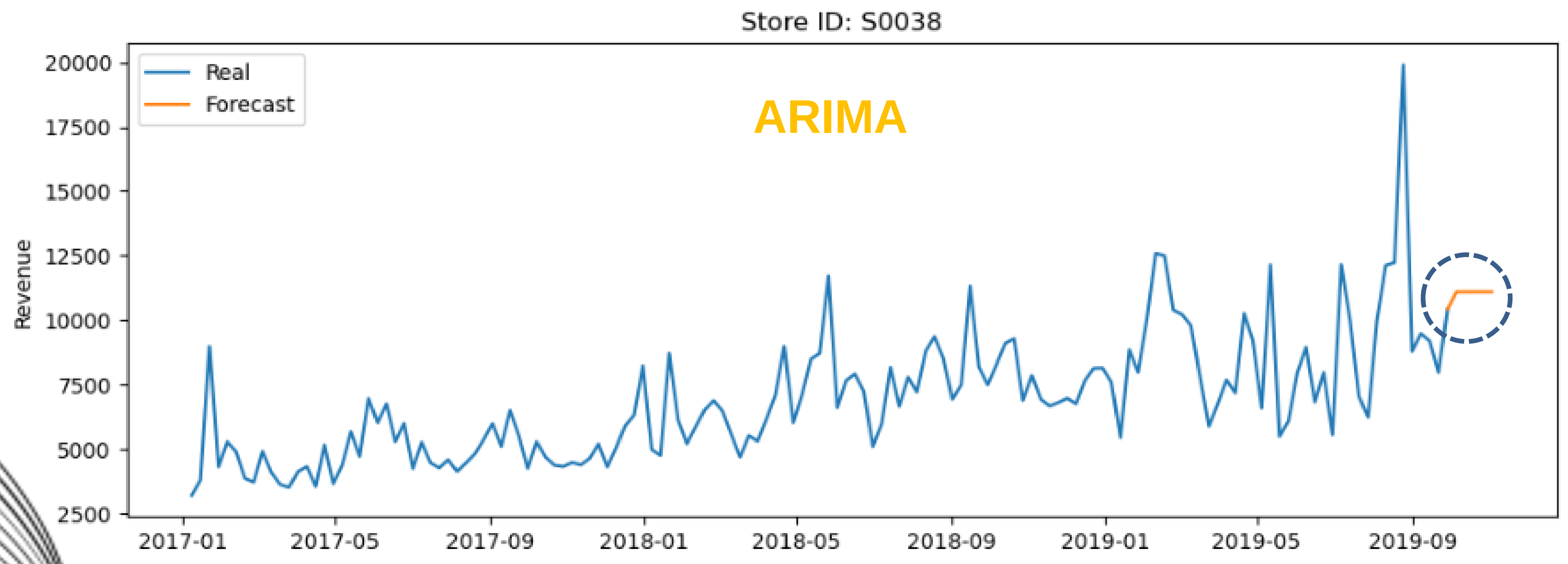


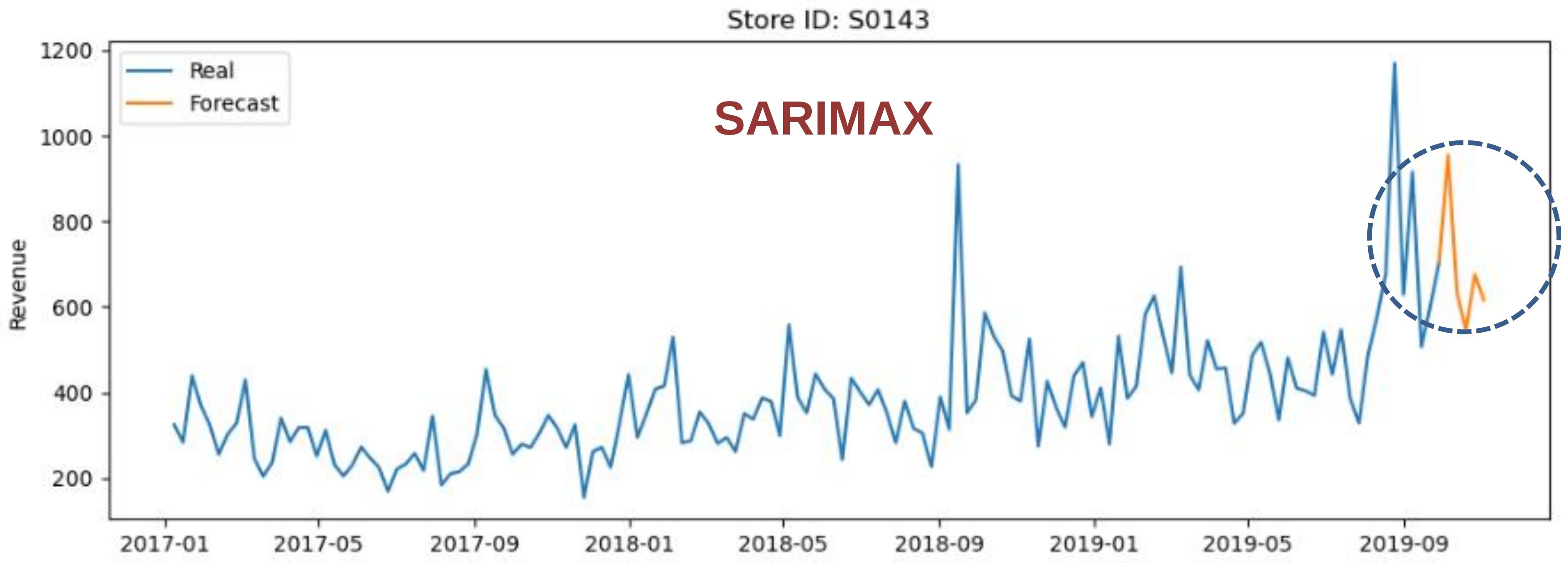
FORECAST

➢ **Previsão do revenue nas 5 semanas seguintes**



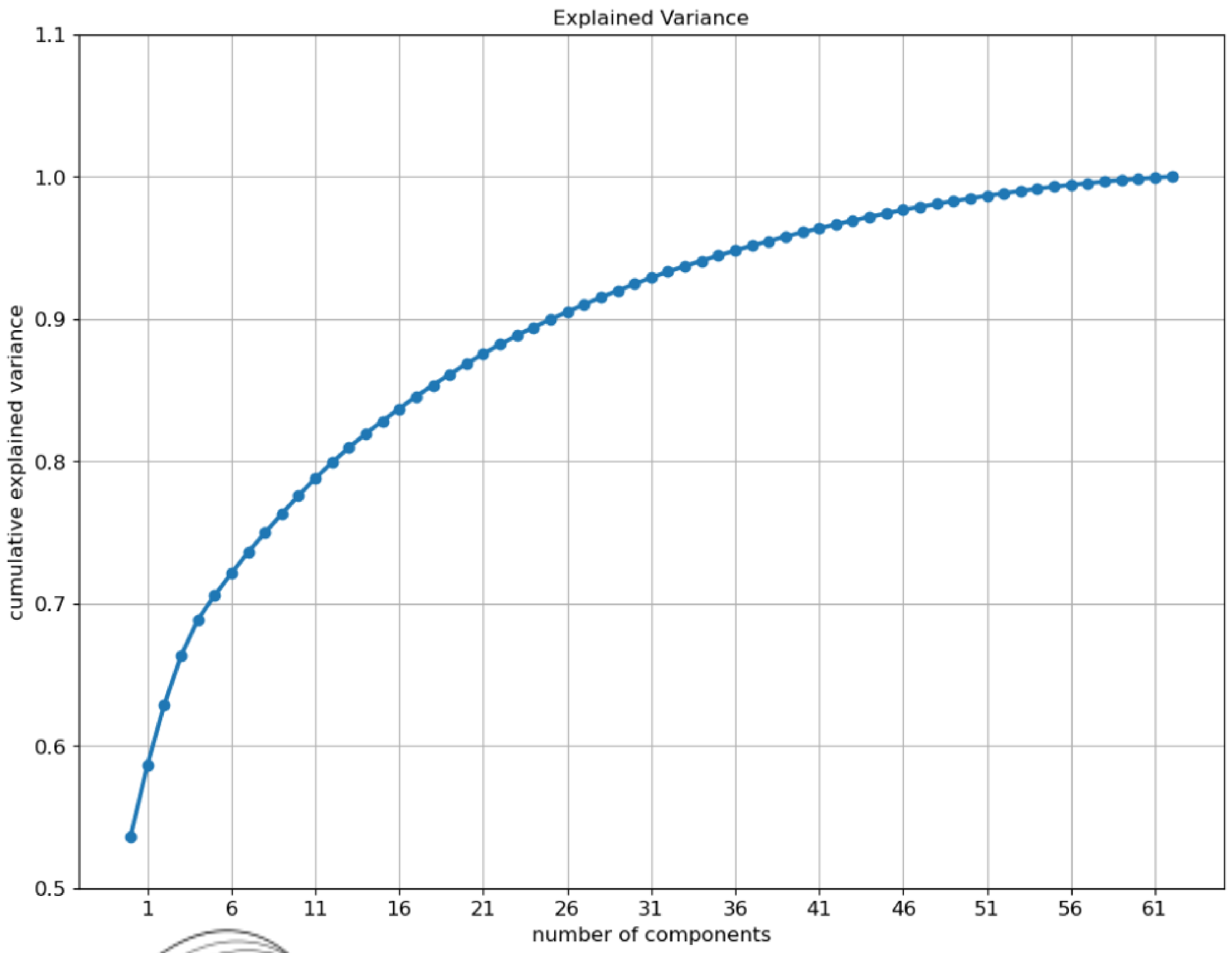


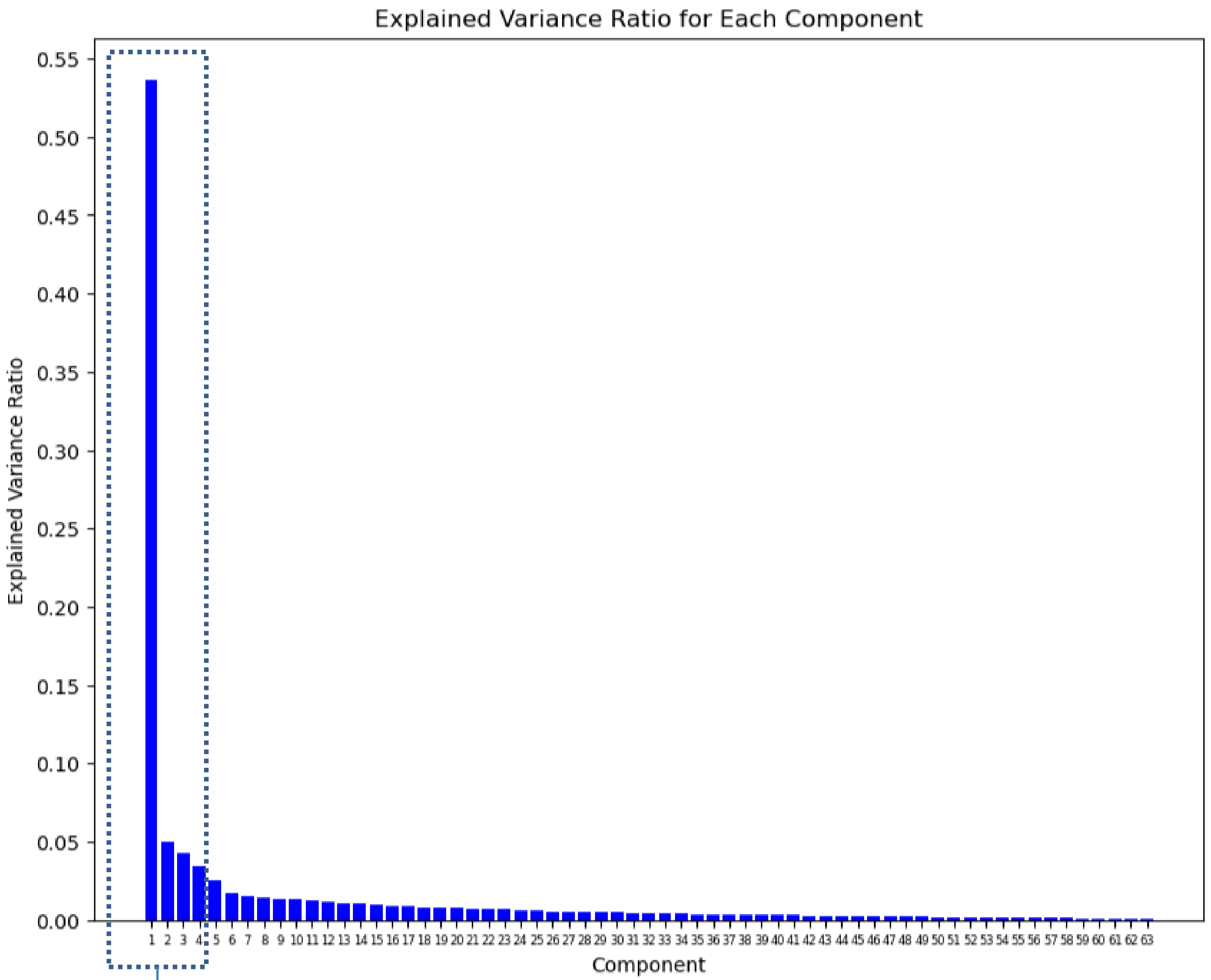




Dimensionality Reduction – PCA

- Objetivo: **Reduzir** o número de lojas em futuras análises
  - Identificar as principais lojas mais contribuem na variância dos dados
- Vantagens:
  - **Reduzir** tempos de processamento
  - **Reduzir** número de máquinas de processamento
  - **Facilitar** a visualização dos dados e a **identificação** de tendências ou padrões



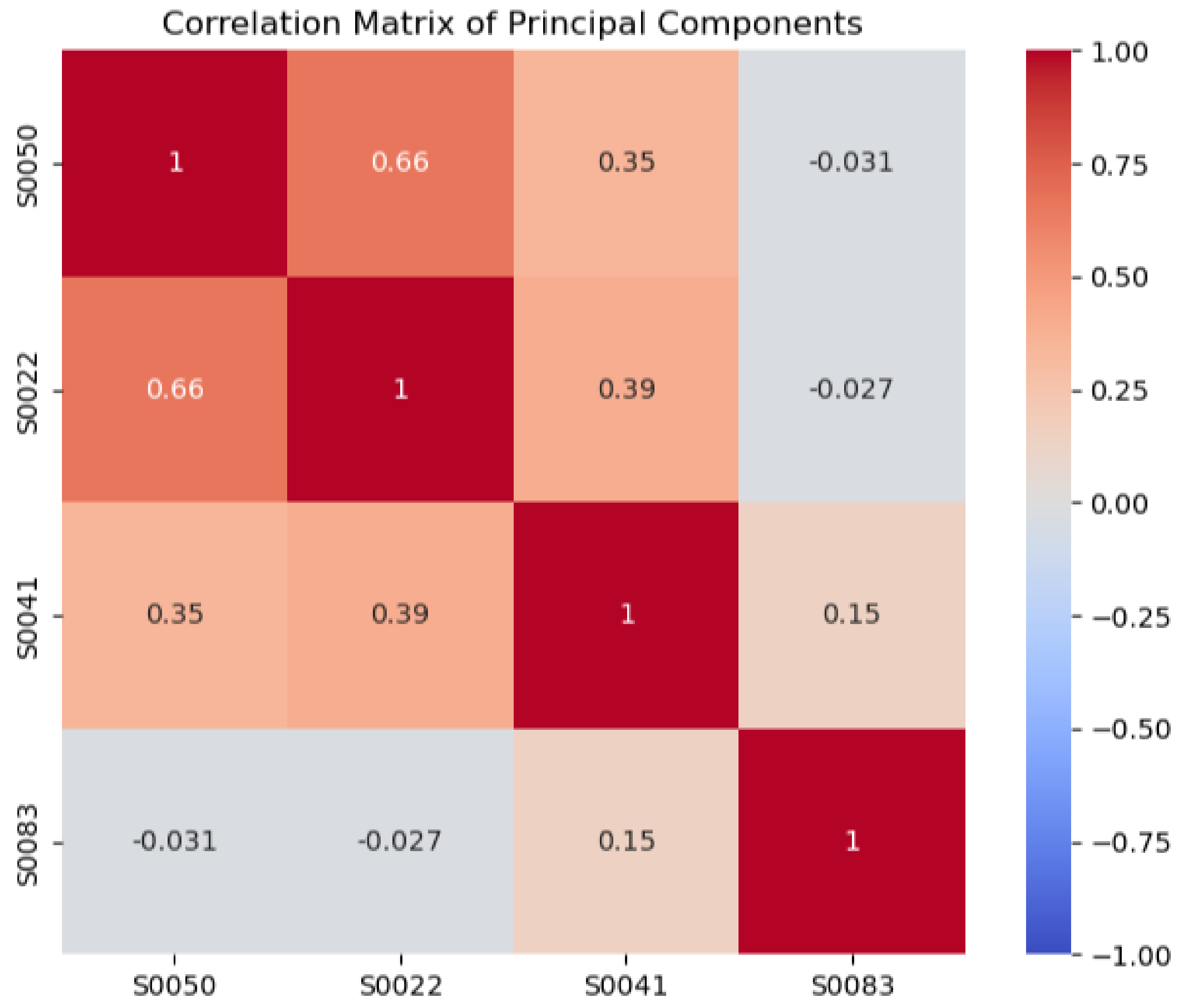


Principal component	Variância explicada	Store ID
PC 1	53.6%	S0050
PC 2	5.0%	S0022
PC 3	4.2%	S0041
PC 4	3.4%	S0083

Total ≈ 66%

Dimensionality Reduction – PCA

Análise de correlação entre os 4 componentes principais

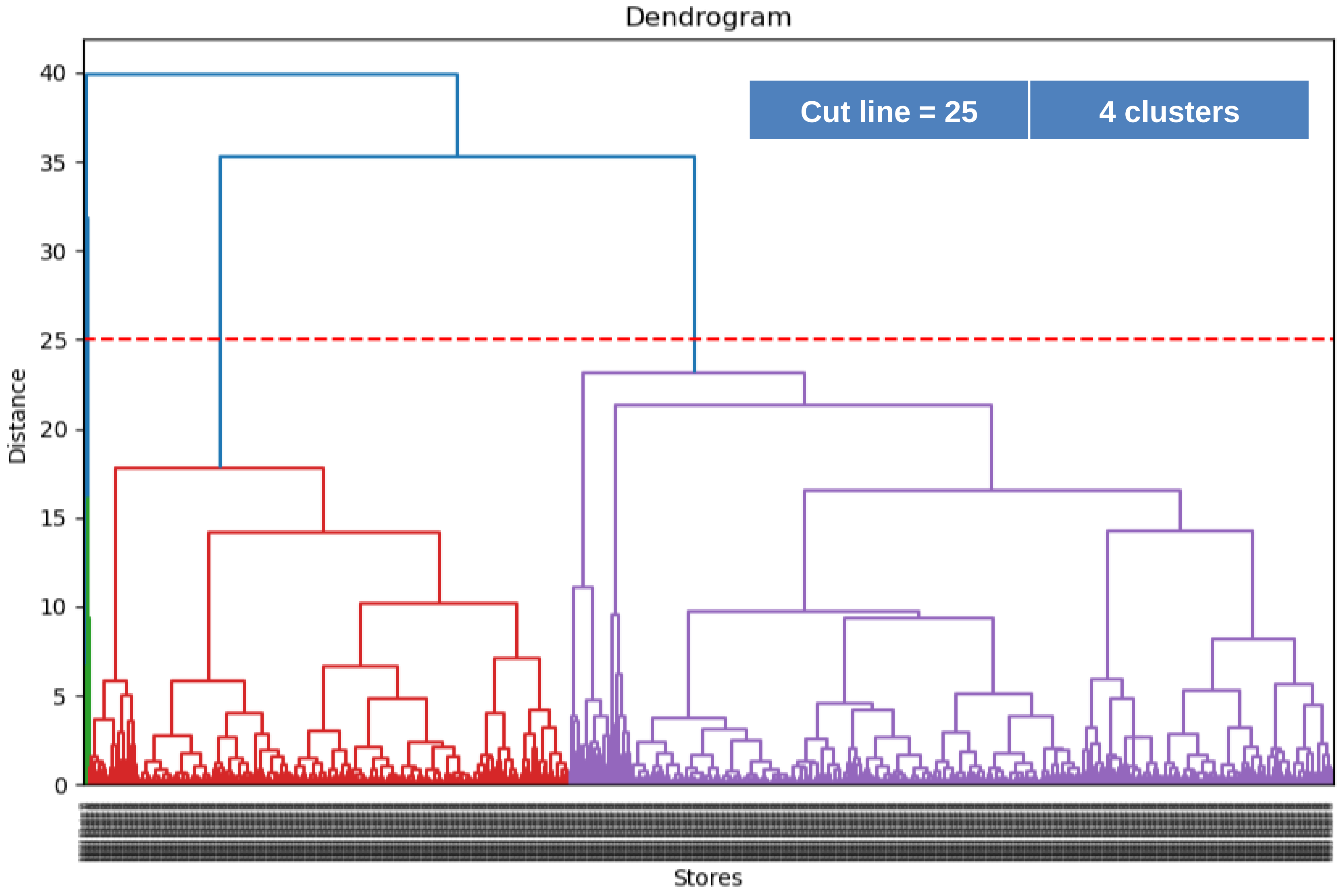


Elevadas correlações entre lojas podem indicar:

- Tendências similares de lucro ao longo do tempo
- Lojas pertencem à mesma cidade
- Lojas são do mesmo tipo

4 PC

Hierarchical clustering (daily revenue)



Hierarchical clustering considera os valores diários do lucro para todas as lojas e identifica grupos de dias com tendências similiares de lucro

Next steps

1. **Melhorar a performance dos modelos de previsão utilizados:**
   - Adicionar outras variáveis exógenas
   - Analisar com mais detalhe a sazonalidade das series temporais

2. **Testar outros modelos de previsão:**
   - Deep learning: RNN / LSTM

3. **Adicionar mais features ao modelo de clustering de forma a detetar outros padrões:**
   - Cidade
   - Tipo de loja
   - Tipos de produtos
   - Número de vendas