

PONTIFÍCIA UNIVERSIDADE CATÓLICA DE MINAS GERAIS
NÚCLEO DE EDUCAÇÃO A DISTÂNCIA
Pós-graduação *Lato Sensu* em Ciência de Dados e Big Data

ALEXANDRE DOS SANTOS FERREIRA

**SÉRIE TEMPORAL NA PREVISÃO DE RECEITAS TRIBUTÁRIAS FEDERAIS E
DO ESTADO DE SÃO PAULO**

Belo Horizonte

2021

ALEXANDRE DOS SANTOS FERREIRA

**SÉRIE TEMPORAL NA PREVISÃO DE RECEITAS TRIBUTÁRIAS FEDERAIS E
DO ESTADO DE SÃO PAULO**

Trabalho de Conclusão de Curso apresentado
ao Curso de Especialização em Ciência de
Dados e Big Data como requisito parcial à
obtenção do título de especialista.

Belo Horizonte

2021

SUMÁRIO

1. Introdução	4
1.1. Contextualização.....	4
1.2. O problema proposto.....	4
2. Coleta de Dados	5
3. Processamento/Tratamento de Dados	7
4. Análise e Exploração dos Dados	9
5. Criação de Modelos de Machine Learning	16
6. Apresentação dos Resultados	24
7. Links	30
REFERÊNCIAS	31

1. Introdução

1.1. Contextualização

O Estado brasileiro é formado por três entidades políticas. Estas entidades são divididas em um Governo Central (União), Governos Regionais (Estados-membros) e Governos Locais (Municípios).

Cada nível de governo tem competência para promover determinadas políticas públicas. E, para realizar essas políticas públicas necessitam arrecadar recursos de diversas maneiras. Sendo que os Tributos são a principal fonte de recursos para que estes governos se mantenham.

Os tributos podem assumir a forma de impostos, taxas e contribuições, estando presentes nas mais diversas atividades.

Como exemplo da aplicação destes tributos temos os serviços públicos (saúde, educação, segurança, e outros) e os mais diversos investimentos (urbanização, saneamento básico, habitação, e outros).

Tendo em vista a importância que estes valores arrecadados representam para os governos, se faz necessário uma certa previsibilidade.

Normalmente os governos fazem planos de curto e de longo prazo, contidos em projetos de lei orçamentária anual ou planos plurianuais. Que, dependem diretamente da arrecadação tributária para serem executados.

Desta maneira, este projeto visa elaborar um modelo preditivo analisando dados passados e analisando diversos aspectos para que haja uma melhor compreensão do futuro. Desta maneira, o modelo pode ser mais uma ferramenta para auxiliar a elaboração de políticas públicas.

1.2. O problema proposto

O problema analisado neste projeto é a Análise Preditiva de Séries Temporais. Conforme contextualizado, a previsão de arrecadação tributária é de extrema importância para que Governos consigam se planejar para o futuro ou executar medidas que venham a sanar falhas nesta projeção.

Serão analisadas duas bases de dados:

- **Receita Federal do Brasil**: Dados de arrecadação tributária federal de todos os entes federativos

- **Governo de São Paulo**: Dados de arrecadação tributária estadual

Tem-se como objetivos desse trabalho:

- Analisar o comportamento das séries temporais ao longo do tempo;
- Analisar de forma comparativa a Arrecadação Tributária Federal, bem como a arrecadação do Estado de São Paulo;
- Criar modelo preditivo nos dois bancos de dados analisados, utilizando duas maneiras diferentes de fazer a predição:
 - Suavização Exponencial Holt-Winters;
 - SARIMA (*Seasonal Autoregressive Integrated Moving Average*).
- Utilizar a linguagem Python dentro da plataforma Jupyter Notebook para realizar as análises.

O período utilizado será de Janeiro de 2004 até Novembro de 2020.

2. Coleta de Dados

Os dados de Receita Tributária Federal, foram extraídos diretamente do site da Receita Federal do Brasil¹. As seguintes observações podem ser realizadas com relação a obtenção dos dados:

- Os dados são fornecidos no formato “ODS”, ou seja, no formato Planilha OpenDocument;
- Os dados são divididos por Ano e por mês, para obter cada um é necessário entrar na página referente ao mês e ano específico;
- Por este motivo, foram necessários extrair 203 diferentes de arquivos, para obter todos os dados;

¹ <https://www.gov.br/receitafederal/pt-br/aceso-a-informacao/dados-abertos/receitadata/arrecadacao/arrecadacao-por-estado>

- Cada planilha contém as seguintes informações:

RECEITAS	AC	AL	AM	AP	BA	CE
IMPOSTO SOBRE IMPORTAÇÃO	75.614	155.284	53.318.171	1.217	83.132.897	41.976.731
IMPOSTO SOBRE EXPORTAÇÃO	-	1.540	(4)	-	532	(44)
IPI - TOTAL	444.176	8.693.992	10.624.704	461.525	148.499.375	36.271.352
IPI - FUMO	-	-	-	-	7.070	-
IPI - BEBIDAS	405.773	1.095.549	3.160.001	20.595	21.920.845	11.132.173
IPI - AUTOMÓVEIS	-	4.641	36.575	3.726	4.869	-
IPI - VINCULADO À IMPORTAÇÃO	44	4.180	6.114.605	-	42.363.285	14.074.246
IPI - OUTROS	38.360	7.589.622	1.313.522	437.204	84.203.307	11.064.933
IMPOSTO SOBRE A RENDA - TOTAL	27.306.995	89.849.955	483.213.068	21.324.460	774.660.369	538.013.097
IRPF	2.560.814	7.806.737	9.743.141	2.678.033	25.462.304	28.174.459
IRPJ	12.755.285	41.568.932	223.793.610	11.543.570	422.232.701	253.595.661
ENTIDADES FINANCEIRAS	158	5.084	2.113.761	48.028	162.959	437.498
DEMAIS EMPRESAS	12.755.127	41.563.848	221.679.849	11.495.541	422.069.742	253.158.163
IMPOSTO S/ RENDA RETIDO NA FONTE	11.990.896	40.473.601	249.676.318	7.102.857	326.965.363	256.242.977
IRRF - RENDIMENTOS DO TRABALHO	9.843.425	34.401.417	115.932.865	6.426.240	240.743.657	192.734.275
IRRF - RENDIMENTOS DO CAPITAL	1.559.909	3.961.844	7.844.602	83.108	52.398.025	36.722.039
IRRF - REMESSAS P/ EXTERIOR	12.506	140.659	120.646.826	26.627	15.707.896	14.125.219
IRRF - OUTROS RENDIMENTOS	575.057	1.969.680	5.252.025	566.881	18.115.785	12.661.443
IMPOSTO S/ OPERAÇÕES FINANCEIRAS	251.451	1.047.842	15.728.265	43.164	3.891.231	4.111.382
IMPOSTO TERRITORIAL RURAL	50.328	182.192	88.546	13.079	1.002.260	320.977
COFINS	14.859.340	51.106.617	344.294.307	11.828.941	465.397.521	339.278.810
FINANCEIRAS	10.448	100.021	190.804	38.500	1.464.050	21.909.356
DEMAIS	14.848.892	51.006.596	344.103.503	11.790.441	463.933.471	317.369.454
CONTRIBUIÇÃO PARA O PIS/PASEP	5.588.639	17.788.642	89.979.736	5.953.406	151.660.505	103.189.246
FINANCEIRAS	1.696	16.441	30.954	6.252	235.935	3.568.616
DEMAIS	5.586.943	17.772.201	89.948.781	5.947.153	151.424.570	99.620.630
CSLL	8.620.764	26.594.427	105.768.231	5.949.894	310.403.412	187.672.971
FINANCEIRAS	1.855	5.467	2.243.712	45.173	223.637	9.103.314
DEMAIS	8.618.909	26.588.961	103.524.519	5.904.722	310.179.775	178.569.658
CIDE-COMBUSTÍVEIS	-	4.642.562	1.996.859	-	10.903.556	2.165.519
CPSSS - Contrib. p/ o Plano de Segurid. Social Serv. Público	6.941.313	21.877.098	22.912.253	21.303.363	60.699.362	44.986.932
OUTRAS RECEITAS ADMINISTRADAS	566.327	2.481.443	21.614.478	(807.023)	2.332.471	9.757.362
SUBTOTAL [A]	64.704.947	224.420.910	1.149.538.614	66.072.025	2.012.583.490	1.307.744.334

- Todas as receitas tributárias federais, exibidas em 32 linhas diferentes e um valor total ao final;
- Todos os estados da federação brasileira, exibidos em 27 linhas diferentes;
- Lembrando que estes dados estão em cada planilha de Mês e Ano distintos. O conjunto destas 203 planilhas está disponibilizado conforme link ao fim do trabalho;
- Estes foram os dados mais precisos encontrados, por este motivo optou-se por baixar todas estas planilhas diferentes.

Os dados da Arrecadação Tributária do Estado de São Paulo, foi obtido diretamente do Portal da Fazenda e Planejamento do Estado de São Paulo². As seguintes observações podem ser realizadas com relação a obtenção dos dados:

- Os dados podem ser baixados no formato “ODS” ou “XLS”. Foi realizada a extração dos dados no formato “XLS”;

² <https://portal.fazenda.sp.gov.br/acessoinformacao/Paginas/Relat%C3%B3rios-da-Receita-Tribut%C3%A1ria.aspx#>

- A Planilha de cada mês já traz a arrecadação histórica nominal, portanto, bastou realizar o download do último mês;
- A planilha contém as seguintes informações:

Ano/Mês	RECEITA TRIBUTÁRIA ^a					
	ICMS	IPVA	AIR ^b	ITCMD ^c	TAXAS ^d	Total
2004	45.415,6	3.951,3	0,0	311,2	1.798,8	51.477,0
jan	3.575,1	1.445,6	0,0	17,6	148,8	5.187,1
fev	3.262,9	753,9	0,0	15,8	122,1	4.154,7
mar	3.469,4	566,5	0,0	42,8	146,9	4.225,6
abr	3.657,0	146,3	0,0	27,4	138,1	3.968,7
mai	3.667,4	133,3	0,0	29,6	152,8	3.983,1
jun	3.831,6	150,3	0,0	31,2	154,1	4.167,1
jul	3.743,0	128,6	0,0	21,2	145,9	4.038,6
ago	3.916,3	125,3	0,0	14,3	157,1	4.213,0
set	4.041,5	140,8	0,0	11,4	139,7	4.333,5
out	4.010,4	107,9	0,0	19,8	146,8	4.284,9
nov	4.130,0	100,7	0,0	23,9	164,0	4.418,5
dez	4.111,0	152,1	0,0	56,4	182,5	4.502,0
2005	50.248,2	4.658,0	0,0	324,6	2.063,9	57.294,7
jan	4.208,9	1.680,9	0,0	11,6	167,6	6.068,9
fev	3.877,7	906,6	0,0	12,1	134,4	4.930,8
mar	3.987,8	676,2	0,0	25,7	173,3	4.863,0
abr	4.165,2	166,4	0,0	21,6	154,6	4.507,8
mai	4.065,2	155,4	0,0	29,4	167,5	4.417,5
jun	4.129,9	186,3	0,0	28,3	178,7	4.523,2
jul	4.103,8	144,1	0,0	26,9	169,6	4.444,4
ago	4.157,1	150,5	0,0	26,9	204,4	4.538,9
set	4.308,0	186,6	0,0	29,8	179,3	4.703,6
out	4.337,2	120,2	0,0	27,2	163,1	4.647,7
nov	4.367,0	119,2	0,0	24,5	175,6	4.686,3
dez	4.540,4	165,8	0,0	60,6	195,8	4.962,5
2006	55.829,9	5.563,3	0,0	403,0	2.350,5	64.146,7
jan	4.500,9	2.037,9	0,0	15,3	198,6	6.752,6
fev	3.695,2	1.042,2	0,0	18,3	149,5	4.905,2

- Trazendo mês a mês a arrecadação tributária por tributo e o total.

Ressalta-se que a intenção é analisar o comportamento de maneira geral. Para isso os valores da Arrecadação Federal se encontram atualizados. Ou seja, são valores correntes. E, os valores do Estado de São Paulo, são valores nominais. Desta maneira será possível analisar cenários distintos e ter certeza de que a modelagem funciona.

3. Processamento/Tratamento de Dados

Inicialmente será apresentado o tratamento realizado nas planilhas, para que fosse possível uma análise. Foi necessária a transformação dos dados extraídos para uma base que reunisse todas as informações de maneira acessível.

3.1 – Dados da Receita Federal do Brasil

Conforme demonstrado acima os dados da Receita Federal estavam distribuídos em 203 arquivos diferentes. Os seguintes passos foram realizados:

1. Identificação dos principais dados:

- a. Este projeto visa identificar um comportamento geral da arrecadação, para tanto não seria necessário ter os dados de cada receita específica;
- b. Somente os dados totais de cada Estado por mês seriam necessários;

2. Após esta identificação, foi realizada uma **extração apenas dos principais dados e gerada uma planilha que contempla a soma da arrecadação de cada Estado em cada mês**. Conforme pode observar uma prévia abaixo:

PERIODO	AC	AL	AM	AP	BA
jan/04	8.196.055	44.664.273	9.819.616	232.926.443	398.719.822
fev/04	6.880.044	29.343.728	198.376.255	6.290.900	317.779.982
mar/04	6.644.264	29.646.976	321.840.650	7.000.692	451.389.711
abr/04	7.932.322	39.141.205	265.230.821	6.980.236	414.844.245
mai/04	7.408.996	30.907.727	420.825.736	10.347.917	443.559.112
jun/04	6.926.111	32.309.999	371.183.156	9.159.299	427.845.564
jul/04	7.993.748	31.860.721	376.588.818	9.162.396	457.411.163
ago/04	7.621.515	29.707.075	272.978.015	18.692.246	465.624.146
set/04	7.241.813	28.407.026	366.952.544	9.108.487	499.995.874
out/04	7.985.406	36.163.264	405.111.467	8.713.008	514.336.118
nov/04	9.097.279	33.185.644	430.990.573	9.413.708	481.457.819
dez/04	7.173.790	26.471.482	512.597.300	6.744.203	599.069.470
jan/05	9.521.437	47.835.830	384.323.508	17.723.339	702.591.346

3. Esta nova planilha conta com os dados dos **27 estados, em 203 meses** seguidos. Totalizando **5.481 registros diferentes**.

3.2 – Dados do Estado de São Paulo

Apesar de os Dados de São Paulo estarem um pouco mais organizados, foi necessário organizá-los melhor antes de iniciar o tratamento dos dados.

Para esta tarefa foram realizadas as seguintes etapas:

- Exclusão das linhas que somavam o total de cada ano;
- Exclusão das linhas anteriores a 2004, para manter a periodicidade igual aos Dados Federais;
- Exclusão da coluna AIR (Adicional do Imposto de Renda), que estava zerada em todas as linhas;
- Optou-se por manter os tributos, para que seja feita análise dos mesmos posteriormente;
- A planilha com as alterações ficou da seguinte maneira:

PERIODO	ICMS	IPVA	ITCMD	TAXAS	total
jan/04	3.575,10	1.445,60	17,6	148,8	5.187,10
fev/04	3.262,90	753,9	15,8	122,1	4.154,70
mar/04	3.469,40	566,5	42,8	146,9	4.225,60
abr/04	3.657,00	146,3	27,4	138,1	3.968,70
mai/04	3.667,40	133,3	29,6	152,8	3.983,10
jun/04	3.831,60	150,3	31,2	154,1	4.167,10
jul/04	3.743,00	128,6	21,2	145,9	4.038,60
ago/04	3.916,30	125,3	14,3	157,1	4.213,00
set/04	4.041,50	140,8	11,4	139,7	4.333,50
out/04	4.010,40	107,9	19,8	146,8	4.284,90
nov/04	4.130,00	100,7	23,9	164	4.418,50
dez/04	4.111,00	152,1	56,4	182,5	4.502,00
jan/05	4.208,90	1.680,90	11,6	167,6	6.068,90
fev/05	3.877,70	906,6	12,1	134,4	4.930,80

- **Contendo 1015 dados**, nas 203 linhas e 5 colunas apresentadas.

4. Análise e Exploração dos Dados

Durante a análise será utilizada a linguagem de programação Python e serão apresentadas as telas do Jupyter Notebook de maneira a facilitar as análises. Os Notebooks completos estão disponibilizados no link ao fim do trabalho.

4.1 – Arrecadação Tributária Federal

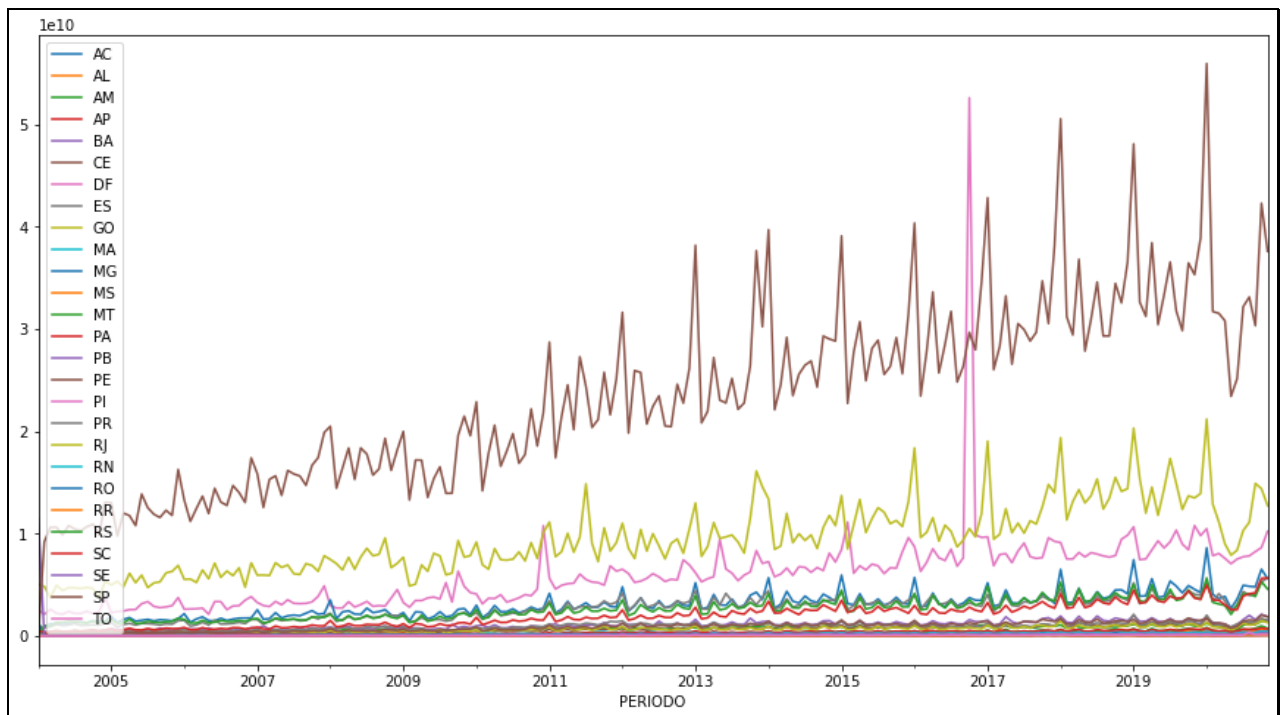
Após a importação da planilha transformada, no formato Excel (.xlsx), foi gerado um DataFrame contendo as seguintes informações:

```
receita_estados.info()

<class 'pandas.core.frame.DataFrame'>
DatetimeIndex: 203 entries, 2004-01-01 to 2020-11-01
Data columns (total 27 columns):
#   Column      Non-Null Count  Dtype
---  -
0    AC          203 non-null    float64
1    AL          203 non-null    float64
2    AM          203 non-null    float64
3    AP          203 non-null    float64
4    BA          203 non-null    float64
5    CE          203 non-null    float64
6    DF          203 non-null    float64
7    ES          203 non-null    float64
8    GO          203 non-null    float64
9    MA          203 non-null    float64
10   MG          203 non-null    float64
11   MS          203 non-null    float64
12   MT          203 non-null    float64
13   PA          203 non-null    float64
14   PB          203 non-null    float64
15   PE          203 non-null    float64
16   PI          203 non-null    float64
17   PR          203 non-null    float64
18   RJ          203 non-null    float64
19   RN          203 non-null    float64
20   RO          203 non-null    float64
21   RR          203 non-null    float64
22   RS          203 non-null    float64
23   SC          203 non-null    float64
24   SE          203 non-null    float64
25   SP          203 non-null    float64
26   TO          203 non-null    float64
dtypes: float64(27)
memory usage: 44.4 KB
```

Observa-se que são 203 entradas (ou seja 203 meses), com 27 colunas (27 estados da federação). Totalizando os 5.481 dados que fundamentarão toda análise.

Os dados plotados se apresentam da seguinte maneira:



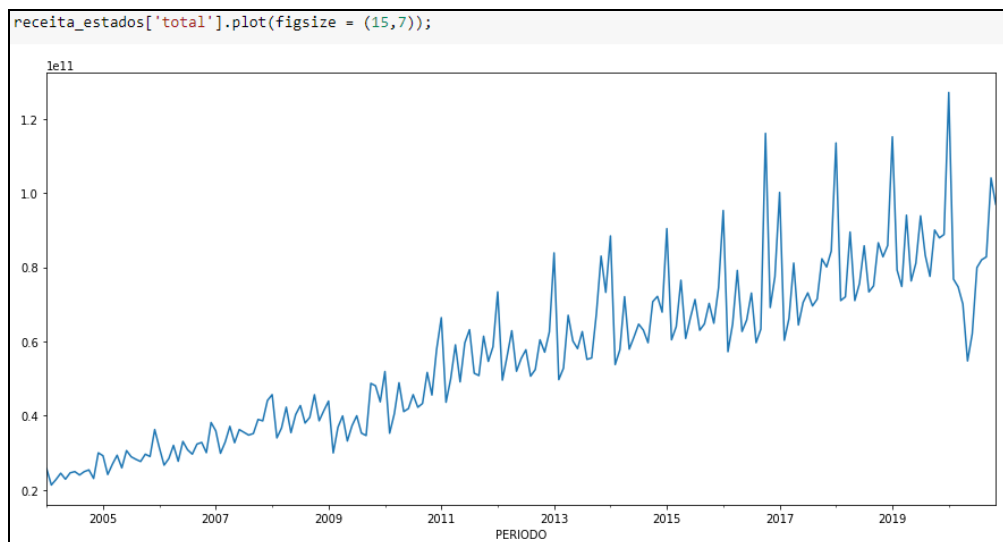
Os Estados com maior arrecadação dos tributos federais são: São Paulo, Rio de Janeiro, Distrito Federal, Minas Gerais e Paraná. Conforme demonstrado abaixo:

```
soma_pd.sort_values(by='Soma', ascending=False).head(5)
```

	Soma
SP	4.736256e+12
RJ	1.913446e+12
DF	1.204153e+12
MG	6.118780e+11
PR	5.399147e+11

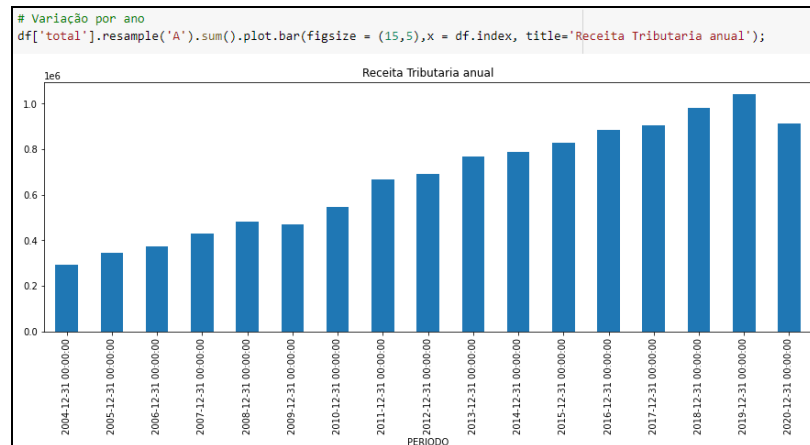
Ao analisar o gráfico é possível observar uma tendência de crescimento nas arrecadações.

A partir deste ponto, foi criada uma coluna com a soma de todas as contribuições de cada Estado (denominada 'total'). Para que seja possível observar a tendência geral da arrecadação. Também serão analisados os principais estados, para checar se o comportamento destes seguem a tendência geral. O gráfico com a soma de todas as arrecadações se apresenta da seguinte maneira:

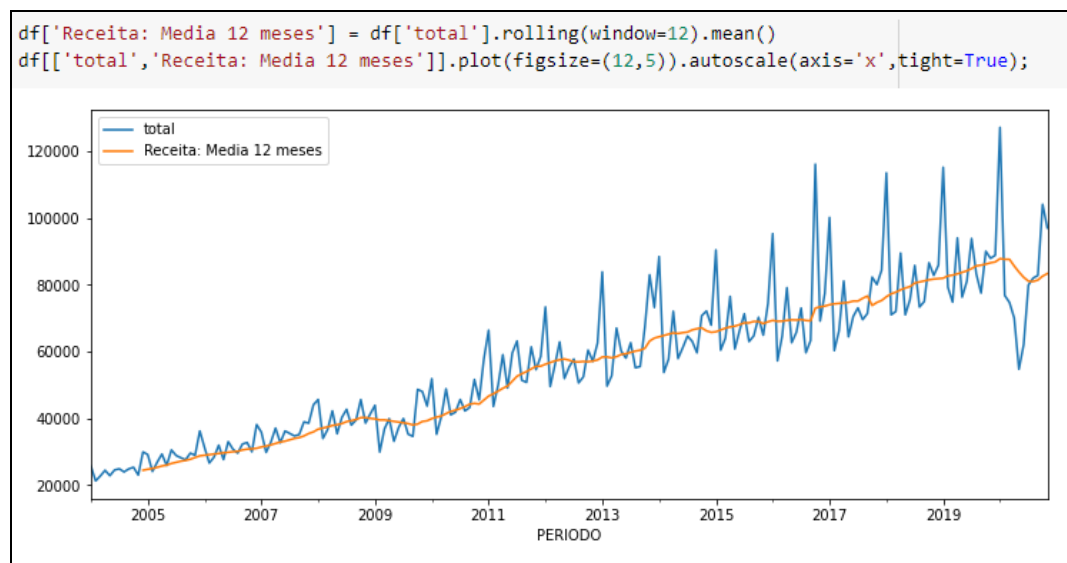


Para facilitar a compreensão dos dados, toda base de dados foi dividida por 1.000.000 (um milhão), para que a partir deste momento os dados sejam apresentados em valores por milhão.

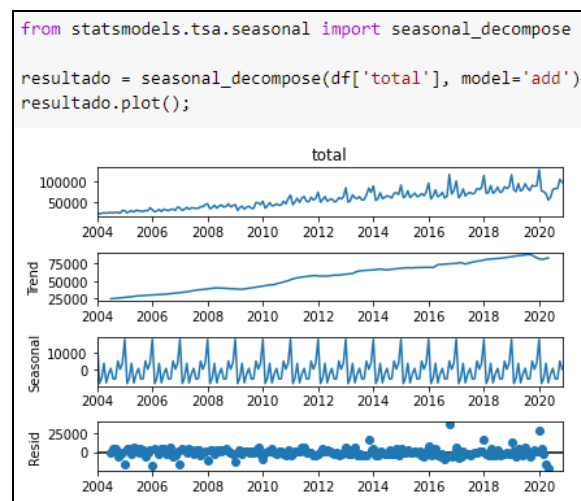
Analisando a receita tributária total por ano podemos observar melhor a tendência de crescimento da receita observada acima:



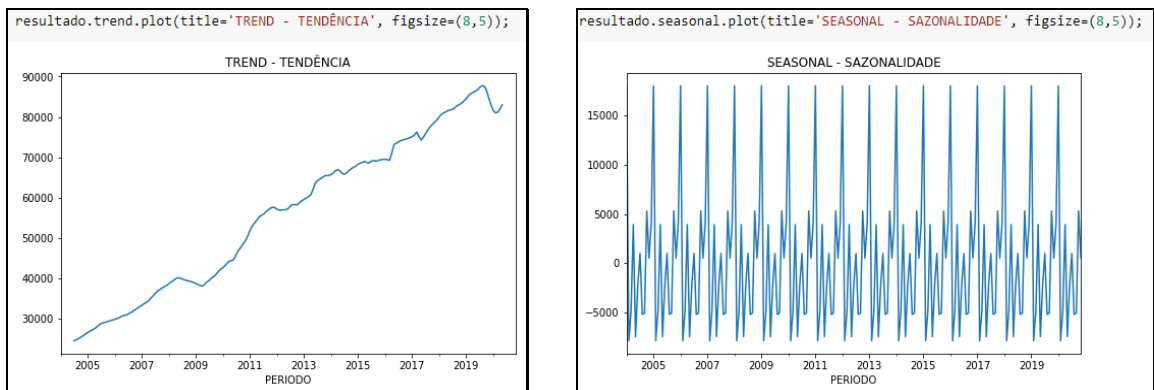
Para facilitar a leitura dos dados e observação de tendência foi inserida uma média de 12 períodos na receita total:



Por fim, será realizada a **decomposição temporal** como forma de isolar componentes individuais de Erro, Tendência e Sazonalidade (ETS).



Ao isolar a componente Tendência e Sazonalidade concluímos tudo o que foi observado até este momento:



Tendo feita toda esta análise dos dados, é possível observar que os dados apresentam certa tendência e sazonalidade. O que já auxilia na escolha de métodos preditivos que melhor se adequem.

Antes de prosseguir com a modelagem, seguem os dados do Estado de São Paulo, que vão permitir a observação se a receita tributária estadual possui um comportamento semelhante à receita tributária federal.

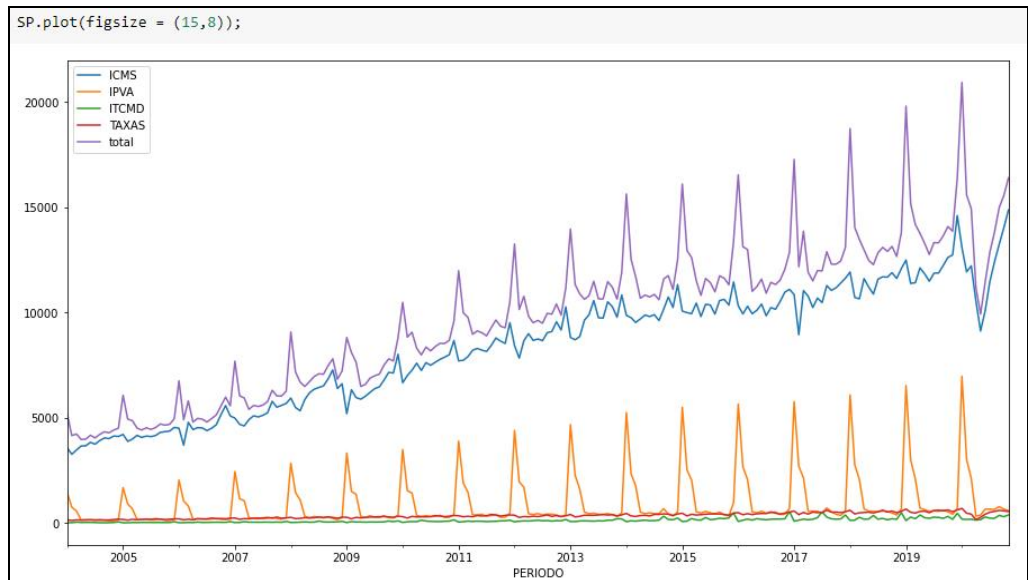
4.2 – Arrecadação Tributária De São Paulo

Após a importação da planilha transformada, em Excel (.xlsx), foi gerado um DataFrame contendo as seguintes informações:

```
SP.info()

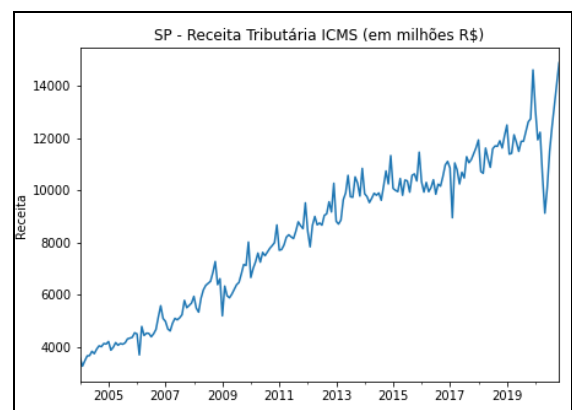
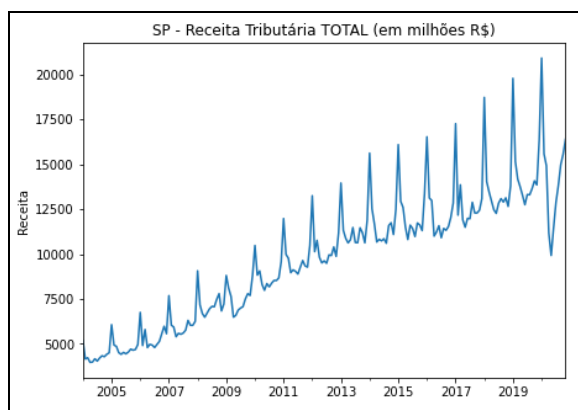
<class 'pandas.core.frame.DataFrame'>
DatetimeIndex: 203 entries, 2004-01-01 to 2020-11-01
Data columns (total 5 columns):
#   Column   Non-Null Count  Dtype
---  ---
0    ICMS     203 non-null    float64
1    IPVA     203 non-null    float64
2    ITCMD    203 non-null    float64
3    TAXAS    203 non-null    float64
4    total    203 non-null    float64
dtypes: float64(5)
memory usage: 9.5 KB
```

Os dados podem ser plotados da seguinte maneira:

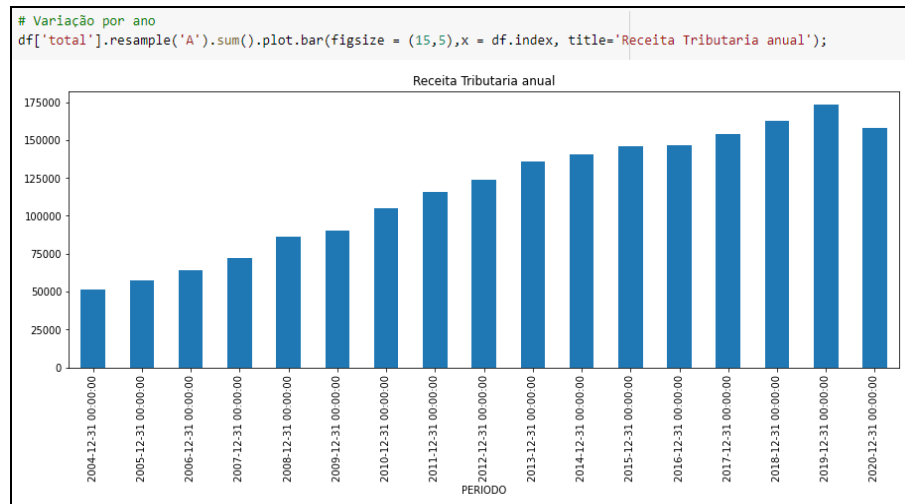


Observa-se que grande parte da arrecadação tributária de São Paulo vem do ICMS (Imposto sobre Operações relativas à Circulação de Mercadorias e sobre a prestação de Serviços de Transporte Interestadual e Intermunicipal e de Comunicação). E, O IPVA tem seu pico no início de cada ano.

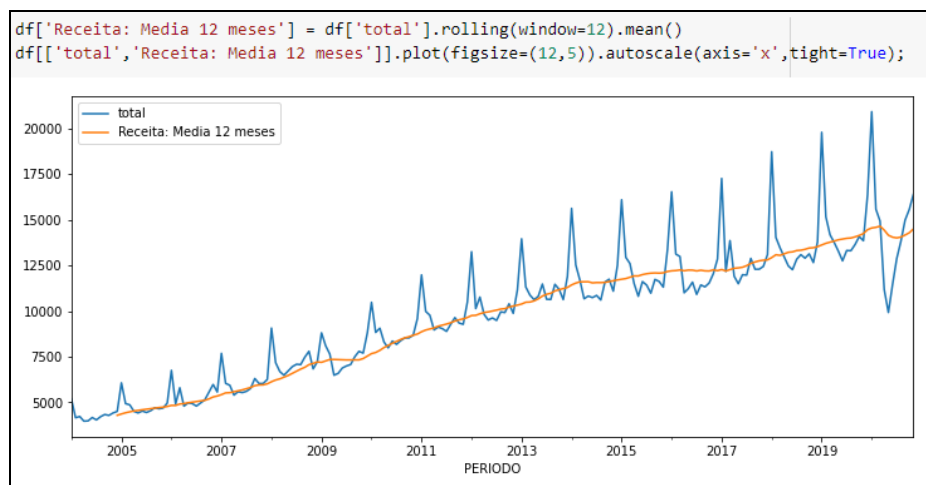
Por este motivo, será analisada a tendência total do Estado de São Paulo e a do ICMS, devido a sua importância. Os gráficos se apresentam da seguinte maneira:



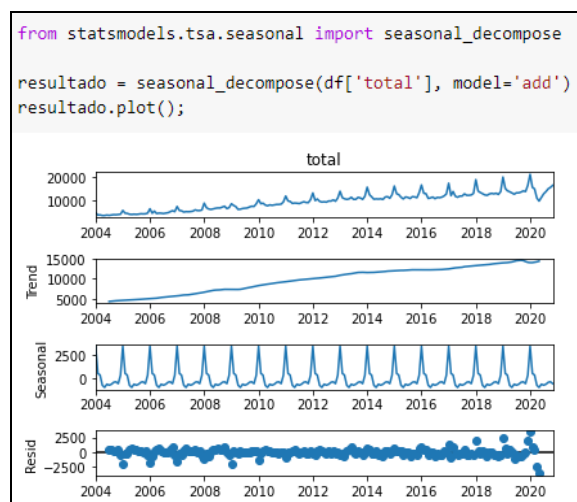
É possível observar uma tendência no aumento da arrecadação, que também pode ser observada na análise de arrecadação anual:



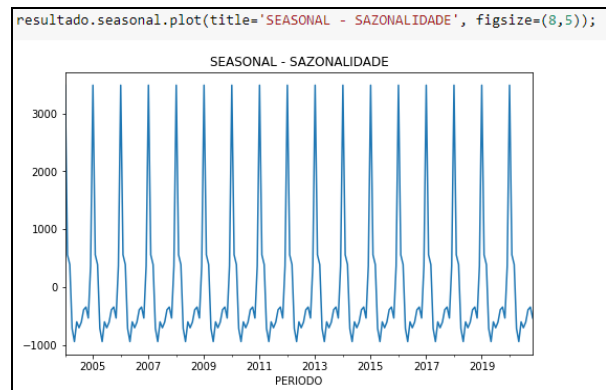
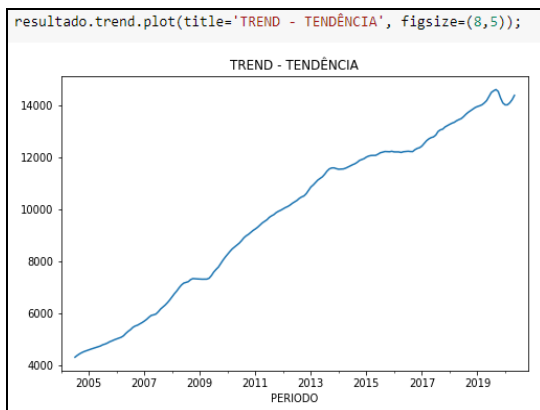
Para facilitar a leitura dos dados e observação de tendência foi inserida uma média de 12 períodos na receita total:



Por fim, será realizada a **decomposição temporal** como forma de isolar componentes individuais de Erro, Tendência e Sazonalidade (ETS).



Ao isolar a componente Tendência e Sazonalidade concluímos tudo o que foi observado até este momento:



A partir da análise e exploração dos dados, serão elaborados modelos que permitam prever comportamentos futuros. Os modelos levarão em conta todas as observações feitas, como tendência, sazonalidade e intervalos de 12 meses.

5. Criação de Modelos de Machine Learning

Para realizar a modelagem serão utilizados 2 métodos: Suavização Exponencial Holt-Winters e SARIMA.

Estes métodos trabalham com séries que apresentam tendência e sazonalidade. Já que tais características foram observadas durante a fase de análise.

Neste projeto será aplicado o método Holt-Winters com sazonalidade aditiva para prever as receitas futuras, desta forma é dada uma importância ao longo do tempo de maneira mais uniforme. A escolha deste método se justifica pelo fato de as previsões realizadas nos métodos exponenciais valorizarem mais os dados mais recentes.

De acordo com Brownlee (2019), o modelo de Média Móvel Integrada Autorregressiva Sazonal, ou SARIMA, é uma abordagem para modelar dados de séries temporais univariados que podem conter tendências e componentes sazonais. Sendo uma abordagem eficaz para a previsão de séries temporais, apesar da exigência cuidadosa na configuração dos hiperparâmetros.

Tendo em vista a complexidade na escolha dos hiperparâmetros, neste projeto optou-se por utilizar a **automatização** na escolha dos mesmos. Será

utilizado o “auto-arima” para que já forneça os parâmetros ideais como ponto de partida das análises.

Os dados serão divididos em bases de treino e teste da seguinte maneira:

- **Treino:** 01/01/2004 até 01/12/2016;
- **Teste:** 01/01/2017 até 01/11/2020.

A divisão foi realizada para chegar o mais próximo de 80% de amostragem para o treino. Mas, de forma que terminasse em dezembro de algum ano.

```
train = df.loc[:'2016-12-01']
test = df.loc['2017-01-01':]
```

5.1 Modelagem das Receitas Federais

5.1.1 Suavização Exponencial Holt-Winters – Receitas Federais

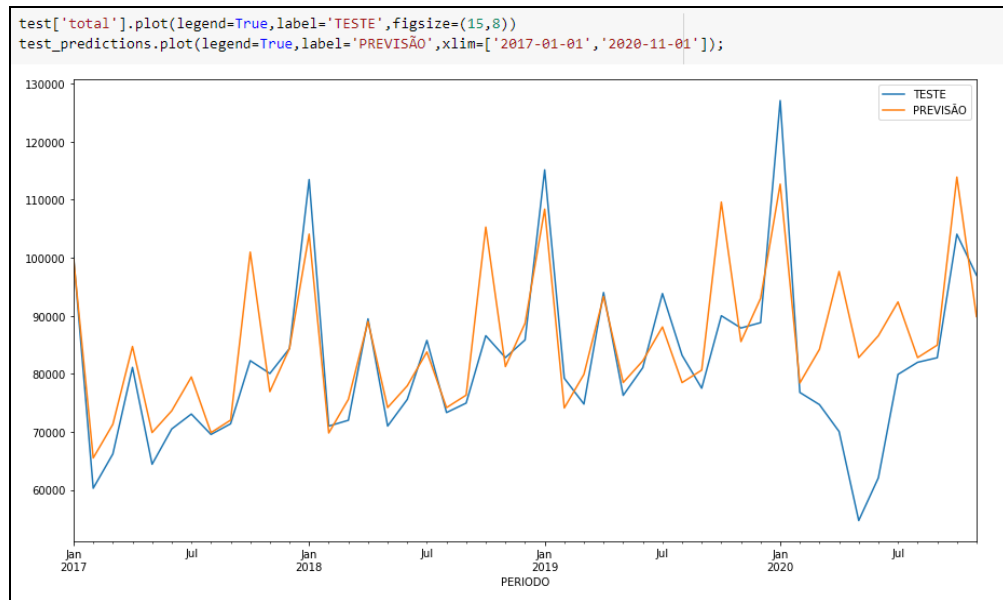
Iniciou-se a modelagem dos dados totais da arrecadação. Sendo aplicado o seguinte código, que contempla as bibliotecas utilizadas:

```
from statsmodels.tsa.holtwinters import ExponentialSmoothing

fitted_model = ExponentialSmoothing(train['total'],trend='add',seasonal='add',seasonal_periods=12).fit()

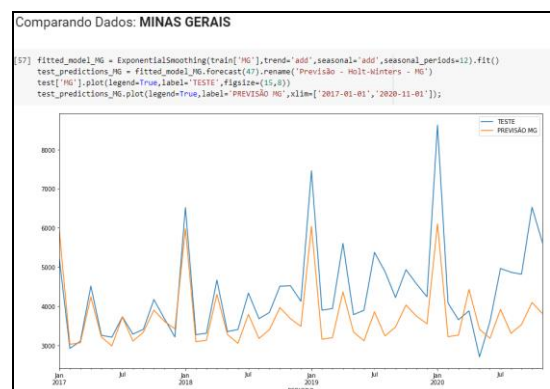
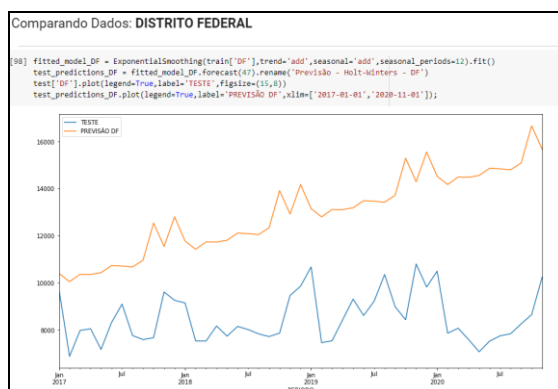
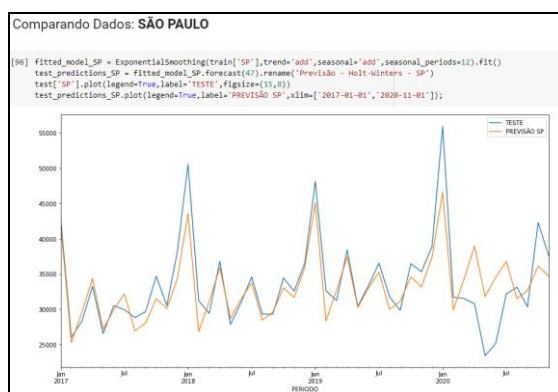
test_predictions = fitted_model.forecast(47).rename('Previsão - Holt-Winters')
```

Ao comparar a base de teste com a previsão feita a partir da base de treino, temos o seguinte resultado:



Como se pode observar o modelo obteve um bom desempenho, conseguindo acompanhar bem a base de teste. A maior parte dos erros se deu entre março e junho de 2020, devido à pandemia do Coronavírus que causou grande colapso em toda economia.

Tendo em vista o bom desempenho na Arrecadação Total, foi realizada a modelagem nos 4 principais Estados, para medir a performance das arrecadações federais em cada Estado específico.



O Outlier no Distrito Federal acabou deixando a previsão um pouco acima da base de teste. O correto seria trabalhar de maneira melhor este outlier, mas conforme informado anteriormente, foi realizada a escolha de não trabalhar os dados para chegar o mais próximo da realidade.

5.1.2 SARIMA – Receitas Federais

Primeiro foi analisada a Receita Tributária Federal Total. Utilizando o “auto-arma”, foram fornecidos os seguintes parâmetros:

```
auto_arima(df['total'],seasonal=True,m=12).summary()
```

SARIMAX Results						
Dep. Variable:	y			No. Observations:	203	
Model:	SARIMAX(1, 1, 1)x(1, 0, 1, 12)			Log Likelihood	-2056.439	
Date:	Tue, 26 Jan 2021			AIC	4122.878	
Time:	19:07:41			BIC	4139.419	
Sample:	0			HQIC	4129.570	
-203						
Covariance Type: opg						
	coef	std err	z	P> z	[0.025	0.975]
ar.L1	0.3053	0.045	6.856	0.000	0.218	0.393
ma.L1	-0.9388	0.025	-37.376	0.000	-0.988	-0.890
ar.S.L12	0.9599	0.021	45.797	0.000	0.919	1.001
ma.S.L12	-0.5574	0.065	-8.641	0.000	-0.684	-0.431
sigma2	3.754e+07	1.48e-09	2.54e+16	0.000	3.75e+07	3.75e+07
Ljung-Box (L1) (Q):	0.28	Jarque-Bera (JB):	1475.99			
Prob(Q):	0.59	Prob(JB):	0.00			
Heteroskedasticity (H):	5.18	Skew:	1.48			
Prob(H) (two-sided):	0.00	Kurtosis:	15.91			

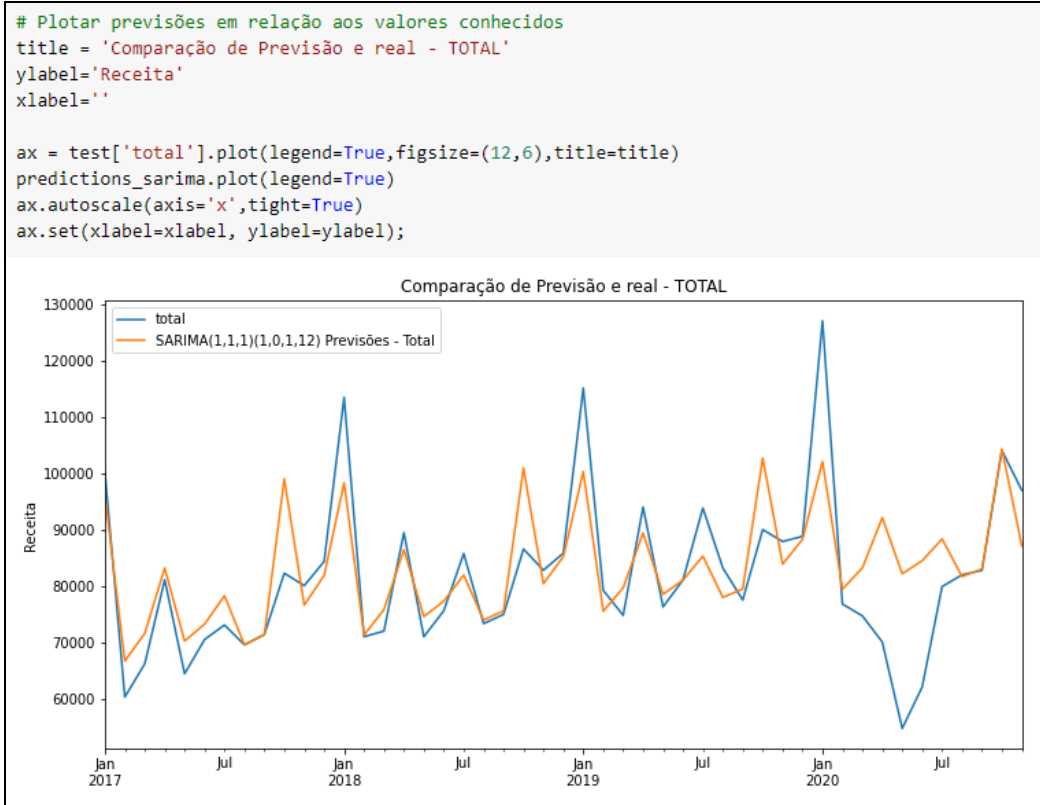
A partir destes parâmetros o modelo foi ajustado e foi feita a previsão da base de teste:

Ajustar modelo SARIMA(1,1,1)(1,0,1,12)

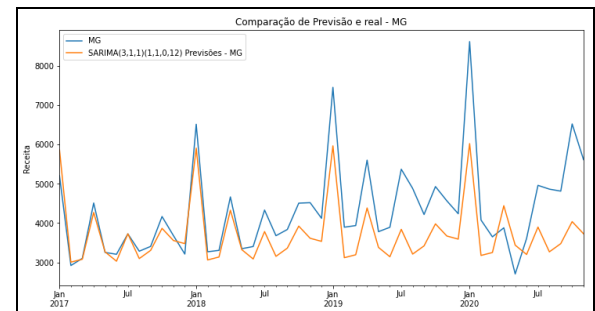
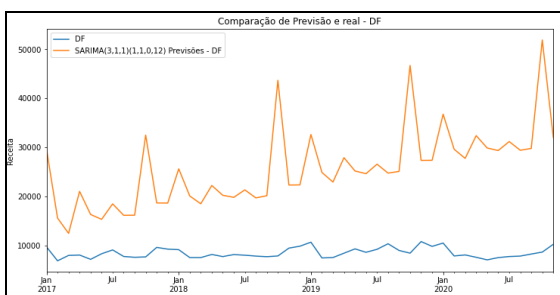
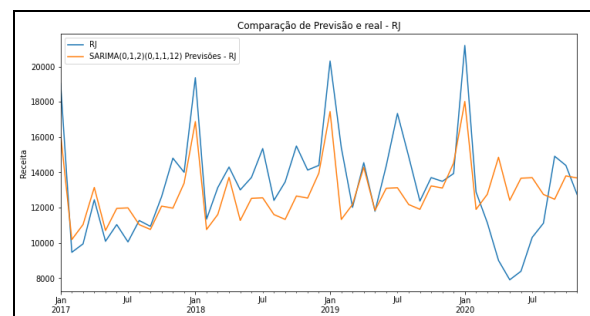
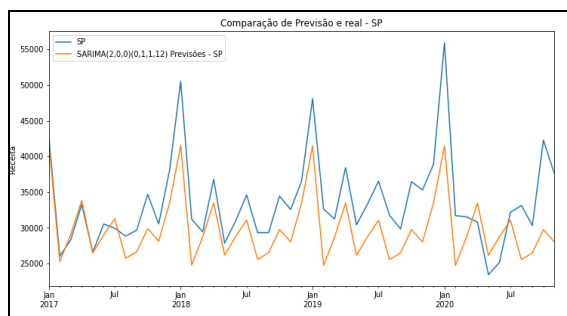
```
model_sarima = SARIMAX(train['total'],order=(1,1,1),seasonal_order=(1,0,1,12))
results_sarima = model_sarima.fit()
results_sarima.summary()
```

```
# Obtendo a previsão
inicio = len(train)
fim = len(train)+len(test)-1
predictions_sarima = results_sarima.predict(start=inicio, end=fim, dynamic=False, typ='levels').rename('SARIMA(1,1,1)(1,0,1,12) Previsões - Total')
```

Após estes passos, foi plotada a previsão, e comparada com a base de teste.



Os mesmos procedimentos foram realizados nos Estados. Cada Estado tem parâmetros próprios que se encontram identificados nos gráficos:



O modelo também apresentou um excelente desempenho na comparação com as bases de teste.

A única observação a ser feita é que ao realizar o auto-arima se fez necessário inserir alguns parâmetros para que fosse encontrada a melhor modelagem o possível. Foi exigido mais processamento, e uma demora um pouco maior na escolha dos parâmetros, mas compensou o esforço empreendido.

Como exemplo destas alterações, segue a configuração para escolha de parâmetros do Rio de Janeiro:

```
auto_arima(df['RJ'], start_p=1, start_q=1,
           max_p=4, max_q=4, m=12,
           seasonal=True,
           d=1,D=1, trace=True,
           error_action='ignore',
           suppress_warnings=True,
           stepwise=False)
```

Estas configurações foram iguais para Distrito Federal e para Minas Gerais. Para o Estado de São Paulo o “auto-arima” simples apresentou bom resultado.

5.2.1 Suavização Exponencial Holt-Winters – Receita Tributária do Estado de São Paulo

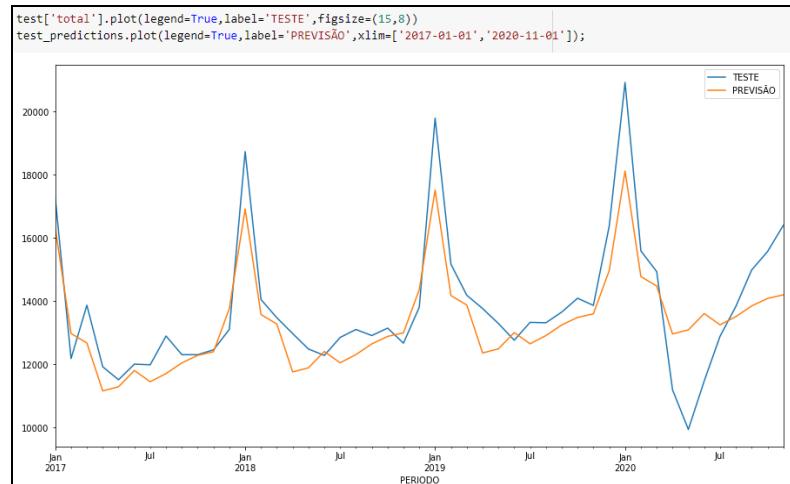
Iniciou-se a modelagem dos dados totais da arrecadação. Sendo aplicado o seguinte código:

```
from statsmodels.tsa.holtwinters import ExponentialSmoothing

fitted_model = ExponentialSmoothing(train['total'],trend='add',seasonal='add',seasonal_periods=12).fit()

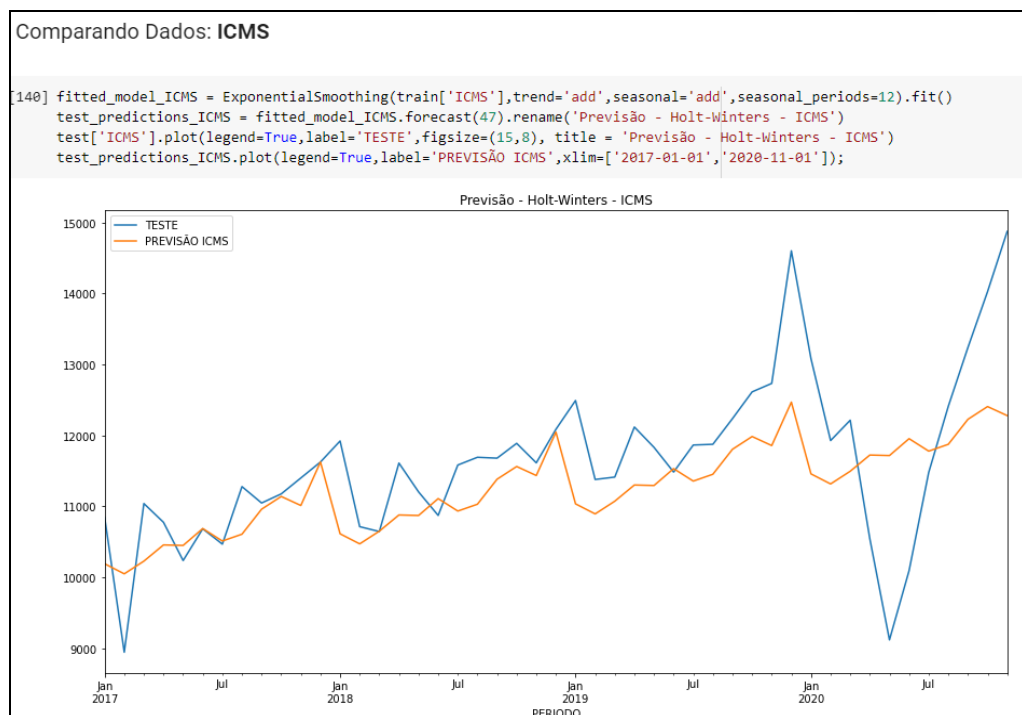
test_predictions = fitted_model.forecast(47).rename('Previsão - Holt-Winters - SP')
```

Ao comparar a base de teste com a previsão feita a partir da base de treino, temos o seguinte resultado:



É possível observar um bom rendimento do modelo em cima da base de testes. Conseguiu acompanhar bem os movimentos e tendência. Sendo possível já observar que o modelo vem funcionando na arrecadação de tributos federais e estadual.

Foi realizada a mesma análise em cima apenas da receita tributária do ICMS. Obtendo o seguinte resultado:



5.1.2 SARIMA – Receita Tributária do Estado de São Paulo

Também foi utilizado o “auto-arima” para auxiliar na definição dos parâmetros. Obtendo o seguinte resultado:

```
auto_arima(df['total'],seasonal=True,m=12).summary()
```

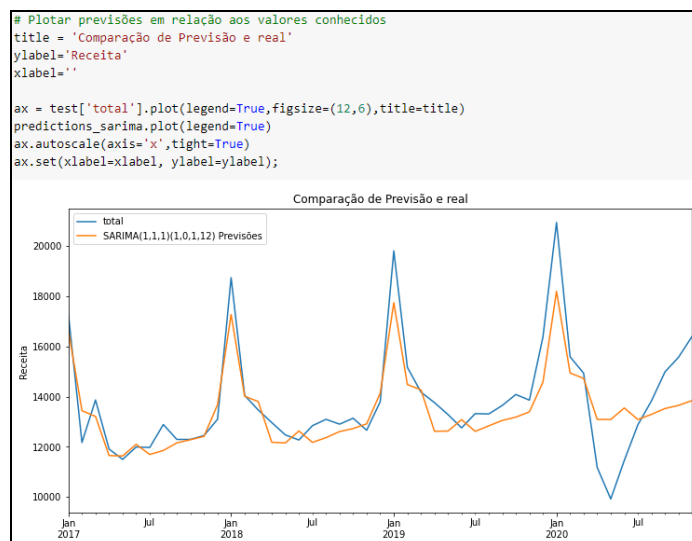
SARIMAX Results						
Dep. Variable:	y			No. Observations:	203	
Model:	SARIMAX(2, 0, 3)x(2, 1, [], 12)	Log Likelihood	-1473.355	AIC	2964.711	
Date:	Tue, 26 Jan 2021	BIC	2993.981	HQIC	2976.567	
Time:	17:11:57					
Sample:	0					
	- 203					
Covariance Type:	opg					
	coef	std err	z	P> z	[0.025	0.975]
intercept	364.8409	134.020	2.722	0.006	102.166	627.515
ar.L1	1.2985	0.238	5.455	0.000	0.832	1.765
ar.L2	-0.6926	0.172	-4.029	0.000	-1.030	-0.356
ma.L1	-0.7895	0.246	-3.209	0.001	-1.272	-0.307
ma.L2	0.3930	0.108	3.630	0.000	0.181	0.605
ma.L3	0.2053	0.121	1.691	0.091	-0.033	0.443
ar.S.L12	-0.2334	0.095	-2.462	0.014	-0.419	-0.048
ar.S.L24	-0.1610	0.093	-1.729	0.084	-0.344	0.022
sigma2	2.871e+05	2.17e+04	13.207	0.000	2.45e+05	3.3e+05
Ljung-Box (L1) (Q):	0.00	Prob(Q):	0.99	Prob(JB):	0.00	
Heteroskedasticity (H):	3.48	Skew:	-1.19	Kurtosis:	8.16	
Prob(H) (two-sided):	0.00					

Realizado o ajuste do modelo e predição da base de testes levando em consideração os parâmetros fornecidos:

```
model_sarima = SARIMAX(train['total'],order=(2,0,3),seasonal_order=(2,1,0,12))
results_sarima = model_sarima.fit()
results_sarima.summary()
```

```
# Obtendo a previsão
inicio = len(train)
fim = len(train)+len(test)-1
predictions_sarima = results_sarima.predict(start=inicio, end=fim, dynamic=False, typ='levels').rename('SARIMA(2,0,3)(2,1,0,12)')
```

Na sequência foi realizada a plotagem dos dados em comparação com a base de dados de teste, já conhecida.



O Modelo SARIMA também apresentou um ótimo desempenho, conforme demonstrado acima.

Agora falta somente extrapolar a base já conhecida e realizar as previsões para períodos futuros.

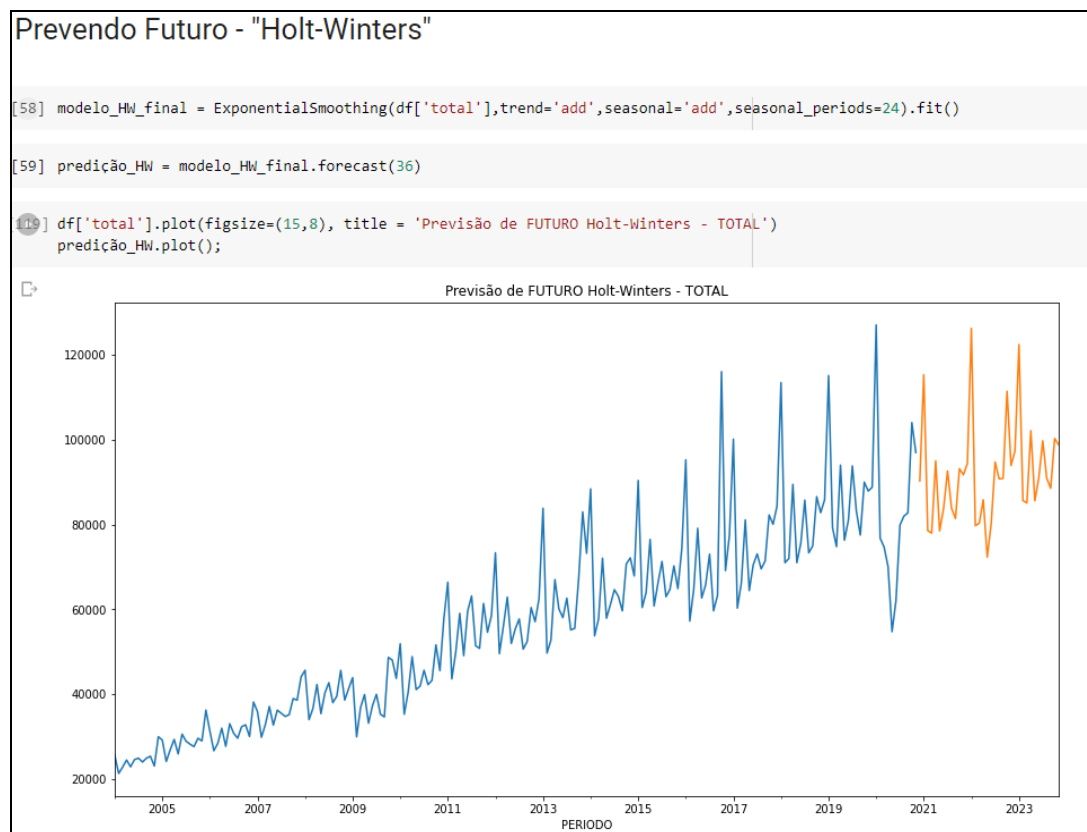
6. Apresentação dos Resultados

Nesta etapa do trabalho serão apresentados os resultados das previsões em tempos futuros. Ressalta-se que quanto mais distante o tempo, mais impreciso o modelo começa a se mostrar. O ideal é sempre reajustar o modelo, para que seja possível aprimorar os cálculos.

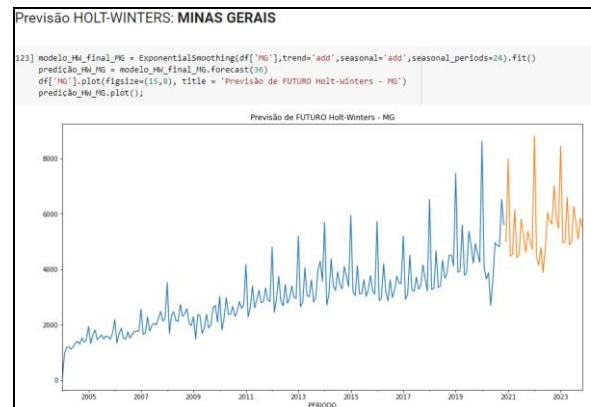
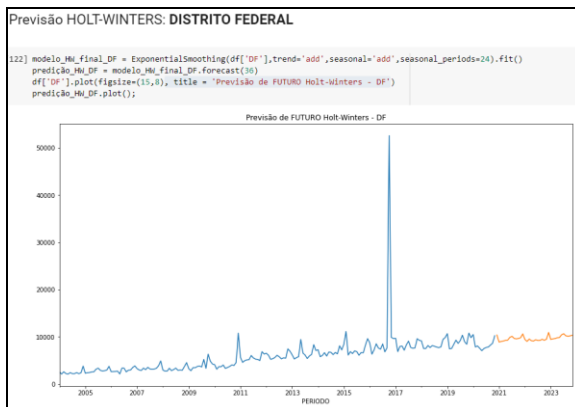
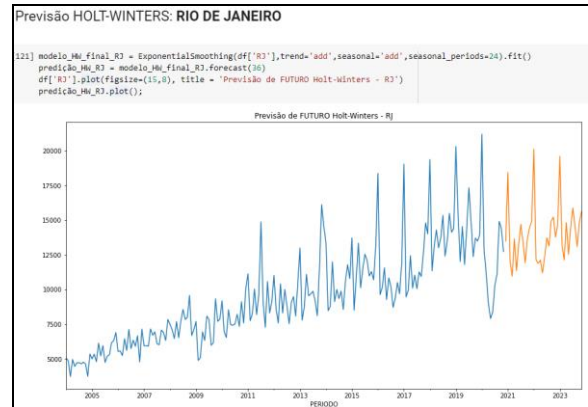
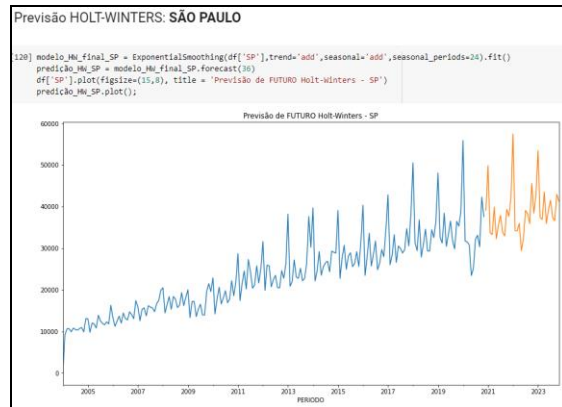
Serão apresentadas as previsões nos dois modelos trabalhados: Holt-Winters e SARIMA. E, nas duas bases de dados trabalhadas. Para que então finalizar com algumas conclusões e insights dos dados apresentados.

6.1.1 Suavização Exponencial Holt-Winters – Receitas Federais

Para fins de melhor visualização foi colocada uma previsão de 36 meses. Mas, em situação real, seria difícil uma previsão tão longa não sofrer ajustes.



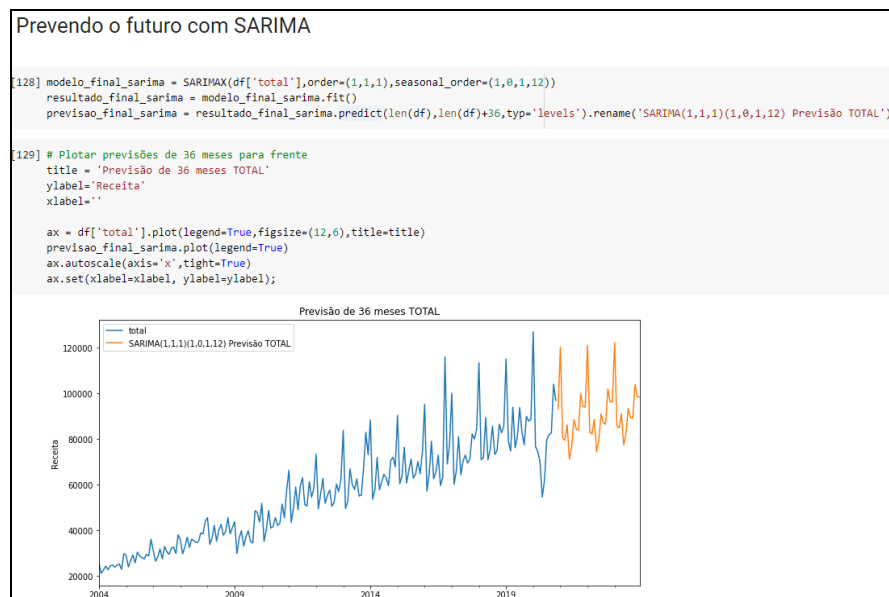
Na sequência segue o desempenho para os 4 Estados estudados até o momento:



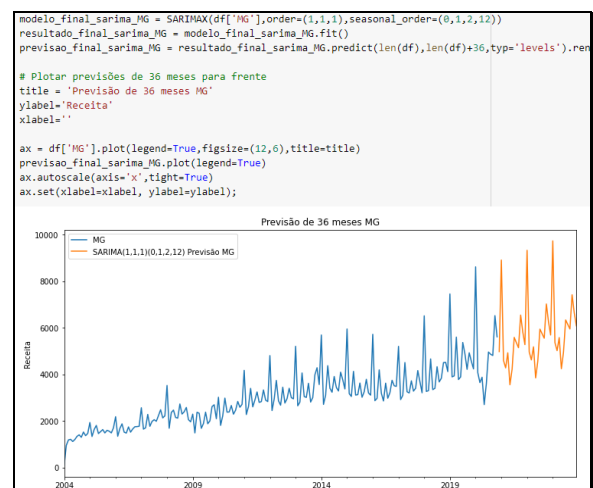
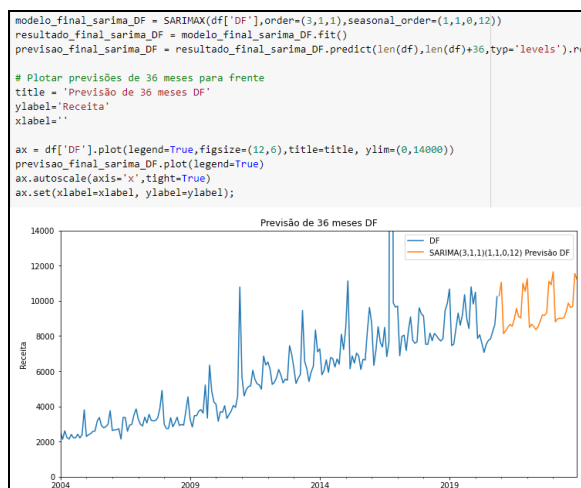
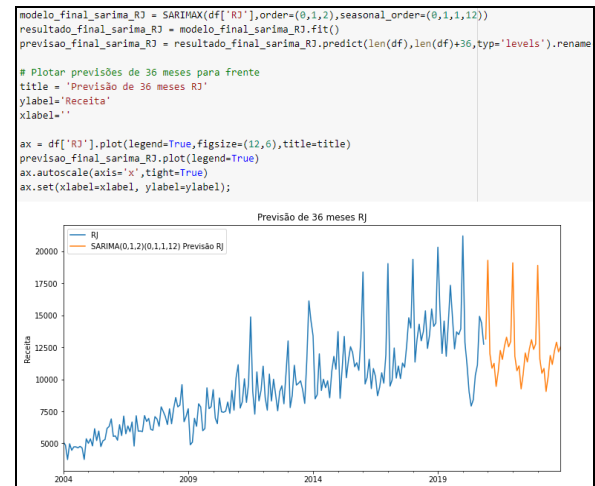
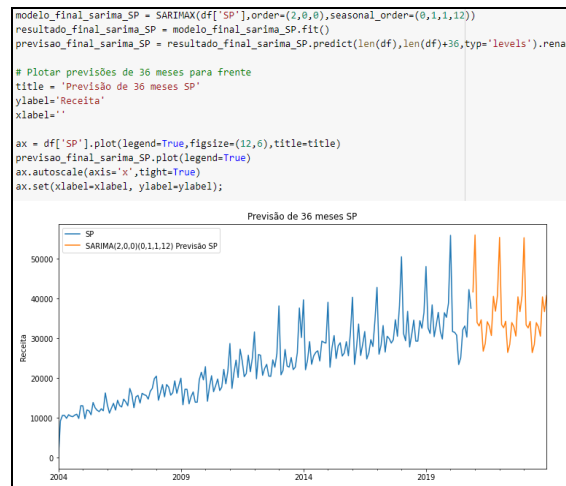
6.1.2 SARIMA – Receitas Federais

Os mesmos parâmetros calculados anteriormente foram mantidos. Será apresentada a previsão de 36 meses.

Segue a previsão da receita total:



Abaixo seguem as previsões dos estados observados:

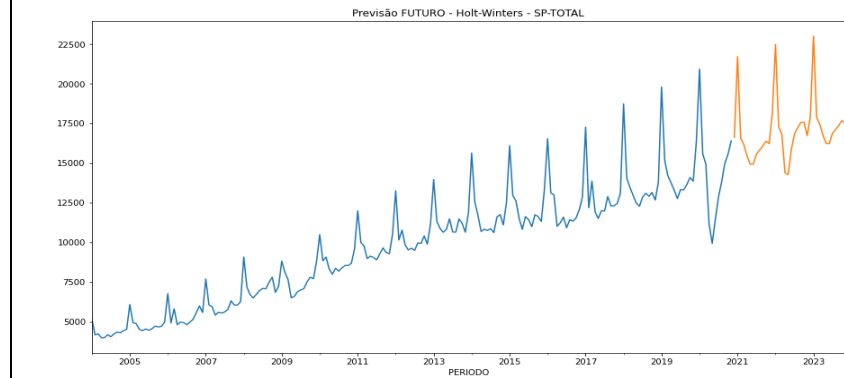


6.2.1 Suavização Exponencial Holt-Winters – Receita Tributária do Estado de São Paulo

Primeiro será apresentada a previsão das receitas tributárias totais e na sequência a previsão do ICMS.

Prevendo Futuro - "Holt-Winters"

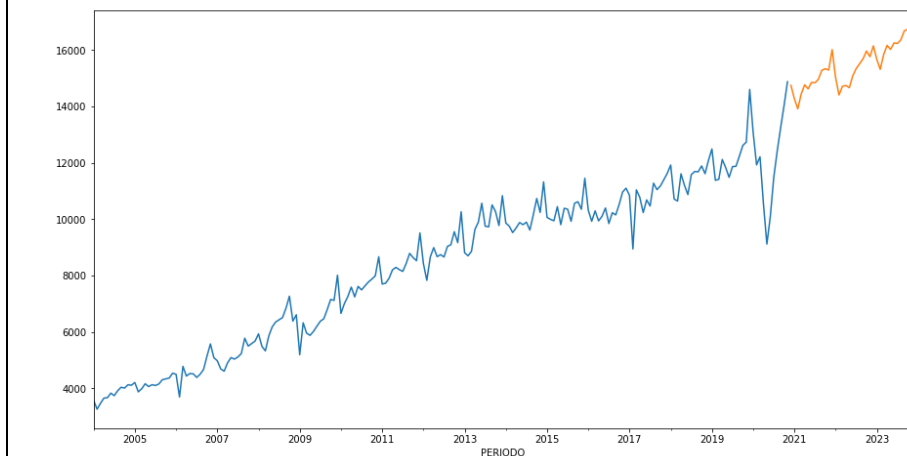
```
[114] modelo_HW_final = ExponentialSmoothing(df['total'],trend='add',seasonal='add',seasonal_periods=24).fit()
[115] predição_HW = modelo_HW_final.forecast(36)
[141] df['total'].plot(figsize=(15,8), title = 'Previsão FUTURO - Holt-Winters - SP-TOTAL')
      predição_HW.plot();
```



Previsão do ICMS:

Previsão HOLT-WINTERS: ICMS

```
[117] modelo_HW_final_ICMS = ExponentialSmoothing(df['ICMS'],trend='add',seasonal='add',seasonal_periods=24).fit()
      predição_HW_ICMS = modelo_HW_final_ICMS.forecast(36)
      df['ICMS'].plot(figsize=(15,8))
      predição_HW_ICMS.plot();
```



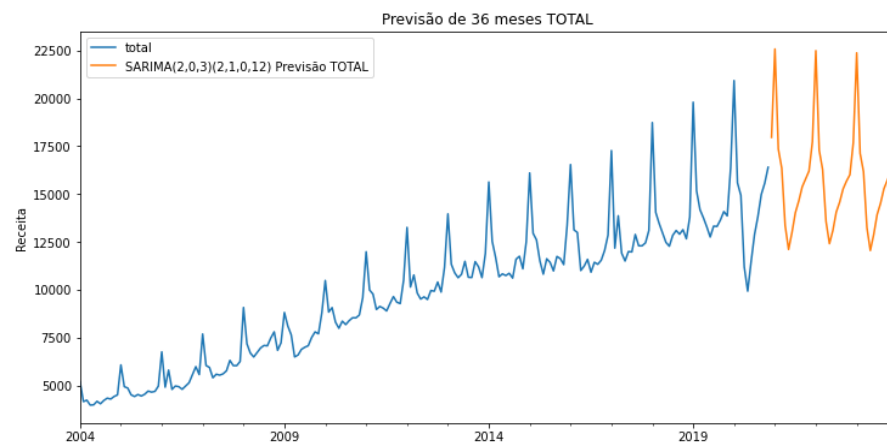
6.2.2 SARIMA – Receita Tributária do Estado de São Paulo

Prevendo o futuro com SARIMA

```
[143] modelo_final_sarima = SARIMAX(df['total'],order=(2,0,3),seasonal_order=(2,1,0,12))
      resultado_final_sarima = modelo_final_sarima.fit()
      previsao_final_sarima = resultado_final_sarima.predict(len(df),len(df)+36,typ='levels').rename('SAR')

[144] # Plotar previsões de 36 meses para frente
      title = 'Previsão de 36 meses TOTAL'
      ylabel='Receita'
      xlabel=''

      ax = df['total'].plot(legend=True,figsize=(12,6),title=title)
      previsao_final_sarima.plot(legend=True)
      ax.autoscale(axis='x',tight=True)
      ax.set(xlabel=xlabel, ylabel=ylabel);
```



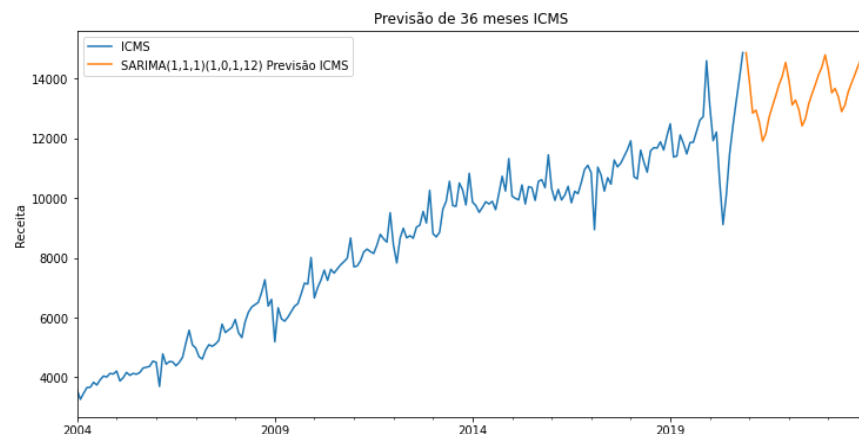
Previsão do ICMS:

Prevendo futuro com SARIMA: ICMS

```
modelo_final_sarima_ICMS = SARIMAX(df['ICMS'],order=(1,1,1),seasonal_order=(1,0,1,12))
resultado_final_sarima_ICMS = modelo_final_sarima_ICMS.fit()
previsao_final_sarima_ICMS = resultado_final_sarima_ICMS.predict(len(df),len(df)+36,typ='levels').rename('SAR')

# Plotar previsões de 36 meses para frente
title = 'Previsão de 36 meses ICMS'
ylabel='Receita'
xlabel=''

ax = df['ICMS'].plot(legend=True,figsize=(12,6),title=title)
previsao_final_sarima_ICMS.plot(legend=True)
ax.autoscale(axis='x',tight=True)
ax.set(xlabel=xlabel, ylabel=ylabel);
```



6.3 CONCLUSÕES

Foi possível observar que os modelos se adaptaram muito bem aos dados, e se saíram bem na base de testes.

Por conta da Pandemia do Coronavírus em 2020, o modelo acabou apresentando mais erros nos meses de grande queda. E, em consequência aprendeu um pouco com essa queda. O ideal para uma maior precisão é fazer ajustes nestes dados que fogem ao padrão.

Sem dúvidas o modelo pode ajudar Governos de maneira geral a prever o quanto deve arrecadar no ano seguinte.

Quando a realidade fugir muito da previsão, políticas podem ser implementadas no intuito de manter aquela estimativa feita no início.

7. Links

O link para o repositório desse trabalho é o seguinte:

Google DRIVE

<https://drive.google.com/drive/folders/10rkUEv4Miw5bs24Y8q7tjbNCmCa0SoDk?usp=sharing>

GITHUB:

https://github.com/alexandre-ferreira-1986/TCC_puc_minas

YOUTUBE:

<https://youtu.be/GCuNTaLwkhU>

Neste diretório constam os seguintes arquivos:

- Bases de dados originais:
 - Planilhas extraídas do site da Receita Federal;
 - Planilha extraída da Fazenda de São Paulo.
- Jupyter Notebooks – 2 arquivos:
 - Um contém a análise da Receita Tributária Federal;
 - Outro contém a análise da Receita Tributária do Estado de São Paulo.
- Bases tratadas: as bases de dados após a transformação dos dados, e que foram utilizadas nas análises;
- Jupyter notebooks em PDF's;
- Este Trabalho de Conclusão em PDF;
- Vídeo de Apresentação (Drive e youtube)

O vídeo de apresentação também foi disponibilizado no Youtube, no seguinte link:

Link para o vídeo: <https://youtu.be/GCuNTaLwkhU>

REFERÊNCIAS

Brownlee, J. (2019). *Deep Learning for Time Series Forecasting*. Machine Learning Mastery.