

BLUE EDTECH

**ANA CRISTINA CHAVES
ANTONIO DUARTE MARCOS JUNIOR**

**BOOTCAMP - PESQUISA DE IMPACTO DO CORONAVÍRUS 2020
SPRINT 3**

**BRASIL
JANEIRO 2023**

1. Introdução

Neste relatório são apresentados os principais resultados obtidos durante a terceira sprint do bootcamp. Para esta sprint os objetivos exigidos foram:

- Evolução da arquitetura dos dados;
- Análise descritiva concluída;
- Correlação dos modelos preditivos;
- Dashboard funcional;
- Material para apresentação intermediária + relatório das atividades (incluindo atividades realizadas e futuras).

Desta forma este relatório foi dividido em **quatro** seções que mostram os resultados de cada um dos itens exigidos.

2. Estruturação dos dados

Nesta seção são apresentados os resultados relativos à primeira tarefa: Iniciar a estruturação dos dados em relação a estruturação e armazenamento.

Foi iniciada a estrutura inicial do data lake seguindo a arquitetura medallion. Na arquitetura medallion os dados são divididos em três diretórios principais: bronze, silver e gold. A seguir são descritos os conteúdos de cada um dos diretórios.

2.1. *Diretório bronze*

No diretório bronze são armazenados os dados na forma mais pura, da forma como foram obtidos originalmente. A estrutura adotada neste diretório é mostrada na Figura 1.

```

+---covid_data
a   +---extract_date=2020-01-22
a   a       CovidData_20200122.csv
a   +---extract_date=2020-01-23
a   a       CovidData_20200123.csv
a   +---extract_date=2020-01-24
a   a       CovidData_20200124.csv
+---ember
a   monthly_full_release_long_format.csv
+---twitter
a   +---query_covid
a   a   +---AR
a   a   a   +---extract_data=2022-12-23
a   a   a   a       CovidTweets_20221223.json
a   a   a   a
a   a   +---CL
a   a   a   +---extract_date=2022-12-25
a   a   a       CovidTweets_20221225.json
a   a   a
a   a   +---EC
a   a   +---ES
a   a   +---MX

```

Figura 1 - Estrutura do diretório bronze.

A seguir é dada uma descrição de cada sub-diretório.

- covid_data
 - Anteriormente este diretório continha apenas os dados fornecidos no dataset original do Kaggle. Porém, foi realizada a coleta de dados a partir do repositório oficial da Universidade de Johns Hopkins. Esta é a mesma utilizada como fonte de dados do dataset original e considerada uma das referências sobre o tema COVID.
 - Dentro deste diretório são criados diretórios que armazenam os dados para cada dia de coleta. Os dados da Johns Hopkins são atualizados diariamente e disponibilizados no formato csv.

- São fornecidos dados, consolidados a partir das fontes oficiais de diversos países, sobre os totais, até a data de coleta, de: casos, mortes, recuperados.
- ember
 - Contém os dados sobre os mercados de energia dos países do projeto obtidos do repositório da EMBER (empresa especializada no uso de dados no mercado de energia a nível mundial).
 - Os dados da EMBER são atualizados mensalmente.
- twitter
 - Conjunto de tweets publicados ao longo da pandemia que contém o termo COVID em seu conteúdo. Como o termo COVID foi criado pela OMS somente a partir de 10 de fevereiro de 2020 esta é a data inicial dos tweets obtidos. Os tweets foram coletados para cada um dos países do projeto.
 - O subdiretório query_covid refere-se a query utilizada para realizar a pesquisa de tweets. No caso foi pesquisado o termo “COVID”.
 - Para cada país é criado um subdiretório a saber: AR (Argentina), CH (Chile), EC (Equador), ES (Espanha), MX (México).
 - Os tweets, por país, são armazenados nos respectivos diretórios com a informação da data em que foram coletados. Em cada coleta são coletados tweets das 24 horas que antecedem.

2.2. *Diretório Silver*

Este diretório contém dados com tratamento prévio. Foram filtrados os dados do diretório bronze referentes apenas aos países de interesse do projeto. Os dados também foram estruturados e organizados para facilitar consultas futuras. A estrutura atual do diretório é exibida na Figura 2.

```

a    ember_electricity_monthly.csv
+---covid_data
a    +---forecast
a    a    +---AR.parquet
a    a    +---CH.parquet
a    a    +---EC.parquet
a    a    +---ES.parquet
a    a    +---MX.parquet
a    +---series_with_calc_fields
a    a    +---AR.parquet
a    +---time_series
a    a    +---AR.parquet
+---twitter
    +---query_covid
    a    +---AR
    a    a    +---tweet_processed
    a    a    a    +---process_date=2022-12-25
    a    a    +---tweet_sentiments
    a    a    +---process_date=2022-12-25
    a    +---CL
    a    +---EC
    a    +---ES
    a    +---MX

```

Figura 2 - Estrutura do diretório silver.

O diretório segue a seguinte estrutura:

- covid_data
 - Contém os dados tratados de COVID. Em todos os subdiretórios os dados são salvos no formato parquet.
 - No subdiretório time_series são salvas as séries temporais de casos de COVID, para os países em estudo, desde o dia 01 de janeiro de 2020.
 - O subdiretório time_series_with_calc_fields contém as séries temporais de casos acrescidas dos seguintes campos

calculados: novos casos, novas mortes, novos recuperados, casos ativos.

- No subdiretório forecast são salvas as previsões de novos casos para cada dia.
- twitter
 - Contém os dados processados dos tweets coletados. Os dados são organizados por query “query_covid” e por país, utilizando a nomenclatura apresentada anteriormente. Neste diretórios os arquivos são salvos no formato json.
 - No subdiretório tweet_processed são salvos os dados tratados dos tweets. A API do Twitter retorna campos que não úteis para a proposta do trabalho, tais campos são removidos e salvos apenas aqueles de interesse. Neste trabalho os campos salvos são: data do tweet, localização e texto do tweet.
 - No subdiretório tweet_sentiments são salvos os resultados da análise de sentimentos dos tweets.
- O arquivo ember_electricity_monthly.csv contém dados mensais do setor de energia elétrica dos países selecionados.

2.3. *Diretório Gold*

Neste diretório são disponibilizados os dados finais que serão utilizados na ferramenta de BI. A estrutura do diretório é apresentada na Figura 3.

```

a    demanda_energia.csv
a
+---covid_data
a    +---forecast
a        +---arima
a            arima_all_countries.csv
a
+---twitter
|    +---sentimental_analysis
|    |    +---AR
|    |    a        process_date=2020-02-11.csv
|    |    +---CL
|    |    +---EC
|    |    +---ES
|    |    +---MX

```

Figura 3 - Estrutura do diretório Gold.

Até o momento os dados são salvos no diretório gold apenas no formato csv por ser um formato de melhor utilização no PowerBi. Nas próximas sprints pretende-se construir um datawarehouse para armazenar os dados finais do projeto.

2.4. Operacionalização dos dados

Para automatizar o processo de coleta e armazenamento dos dados foi iniciado o projeto de uma data pipeline utilizando o Airflow. O Airflow é uma ferramenta que permite o agendamento de tarefas (chamadas de DAG's) a serem executadas em intervalos de tempo pré determinados de forma automática.

A data pipeline está disponibilizada no repositório do Github no diretório datapipeline. No diretório estão os scripts criados para automatizar a coleta e armazenamento dos dados.

No momento desta sprint duas DAG's estão operacionais. A DAG necessária para extração e transformação dos tweets e a para extração, transformação e previsão de casos de COVID

A DAG Twitter realiza a extração dos tweets com o termo covid no texto para cada um dos cinco países de estudo. São coletados 100 tweets diários, são coletados tweets nas 24 horas que antecedem o momento de execução da DAG. Os tweets são salvos na pasta bronze no diretório específico. Em seguida, realiza a transformação do tweets para extração das informações: data de postagem, local de postagem e texto do tweet. O resultado desta tarefa é salvo na pasta silver. Finalizada a tarefa anterior é executada a tarefa de tradução de avaliação de sentimentos dos tweets cujo resultado é salvo em diretório específico.

Na DAG Covid são executadas as tarefas de extração, transformação e carregamento dos dados de Covid. A primeira tarefa faz o download dos dados mais recentes sobre os casos de Covid do repositório da Johns Hopkins. A segunda tarefa realizada é a construção das séries temporais de cada país. O repositório da Johns Hopkins oferece um arquivo para cada dia, nesta tarefa os dados são organizados em um único arquivo no formato parquet. A próxima tarefa executa o cálculo dos seguintes campos: novos casos, novas mortes, novos recuperados, casos ativos. Após esta tarefa é executado o modelo de previsão para gerar a previsão de novos casos para os próximos 7 dias. A última tarefa realiza a carga dos dados no diretório gold.

3. Modelos desenvolvidos

Este projeto exige o desenvolvimento de modelos com objetivos distintos. A princípio foi identificada a necessidade de se desenvolver para realizar duas tarefas distintas: análise de sentimentos e previsão do número de casos de COVID em cada país. Portanto, nesta seção são apresentados os principais resultados dos modelos desenvolvidos para cada uma das tarefas mencionadas.

3.1. Análise de sentimentos

A análise de sentimentos dos tweets foi utilizado o modelo VADER que está disponível na biblioteca NLTK. Antes de se utilizar o VADER da NLTK foi necessário realizar a tradução dos tweets para o inglês. Isto foi necessário pois o VADER é um modelo treinado para textos em inglês.

Para realizar a tradução dos textos, inicialmente, foi utilizado o modelo Marian da biblioteca Spark-NLP, descrito no relatório da sprint anterior. Porém, tal modelo exigia demasiado poder computacional pois, o mesmo foi desenvolvido para utilizar processamento em paralelo em clusters avançados. Para solucionar o problema de tradução dos textos foi utilizada a API do Google Tradutor.

A consulta à API do Google Tradutor foi feita por intermédio da biblioteca Translators. Esta biblioteca disponibiliza métodos para realizar a consulta a diferentes serviços de tradução de texto online como: Google Tradutor e Bing

Tradutor. O único inconveniente é que ela demanda conexão com a internet e o tempo de resposta com o servidor da Google depende da velocidade de conexão, o que torna o processo lento. Porém, esta alternativa ainda teve melhor desempenho de velocidade do que a alternativa utilizando o Spark-NLP.

Com os tweets pôde-se então realizar a análise de sentimentos utilizando o VADER do NLTK. O VADER (Valence Aware Dictionary and Sentiment Reasoner) é uma ferramenta de análise de sentimentos baseada em léxico e regras que está especificamente sintonizada com os sentimentos expressos nas mídias sociais. Por este motivo é o ideal para uso neste projeto.

3.1.1. Resultados

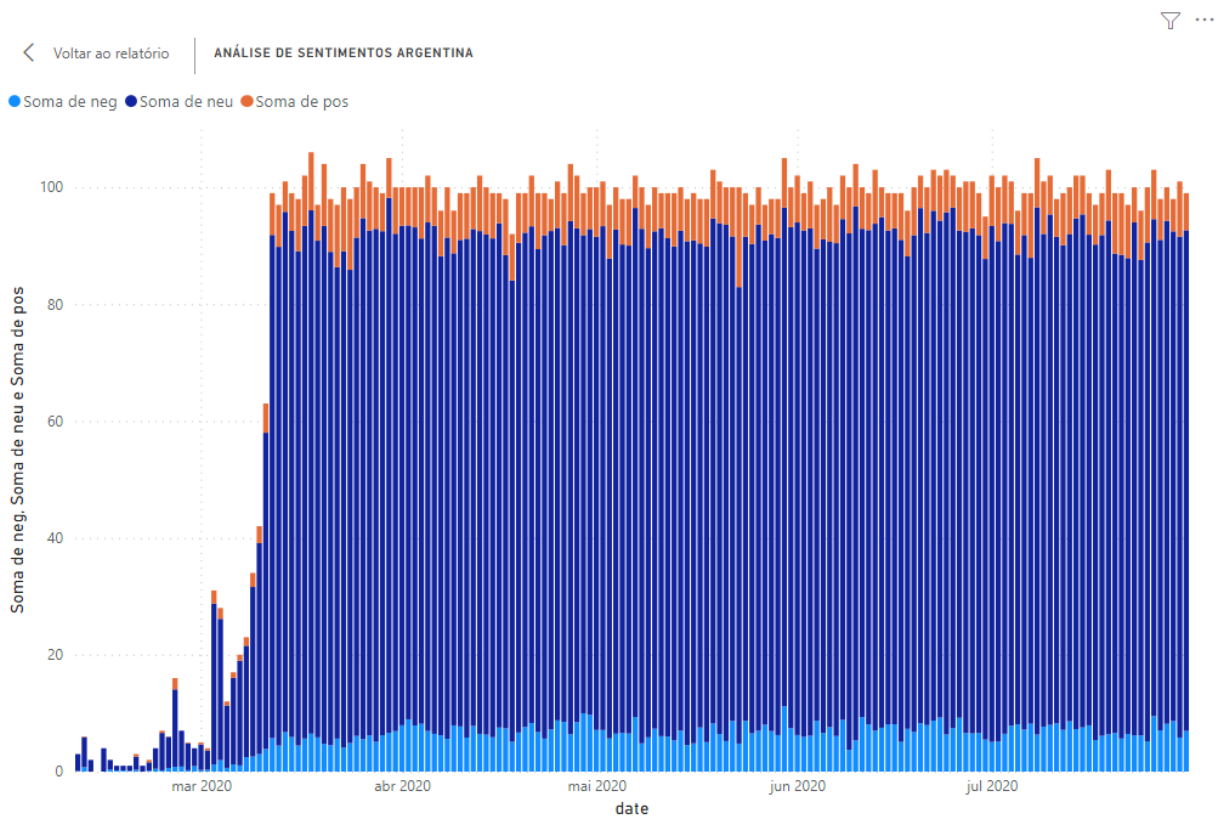


Figura 4 - Análise de sentimentos Argentina

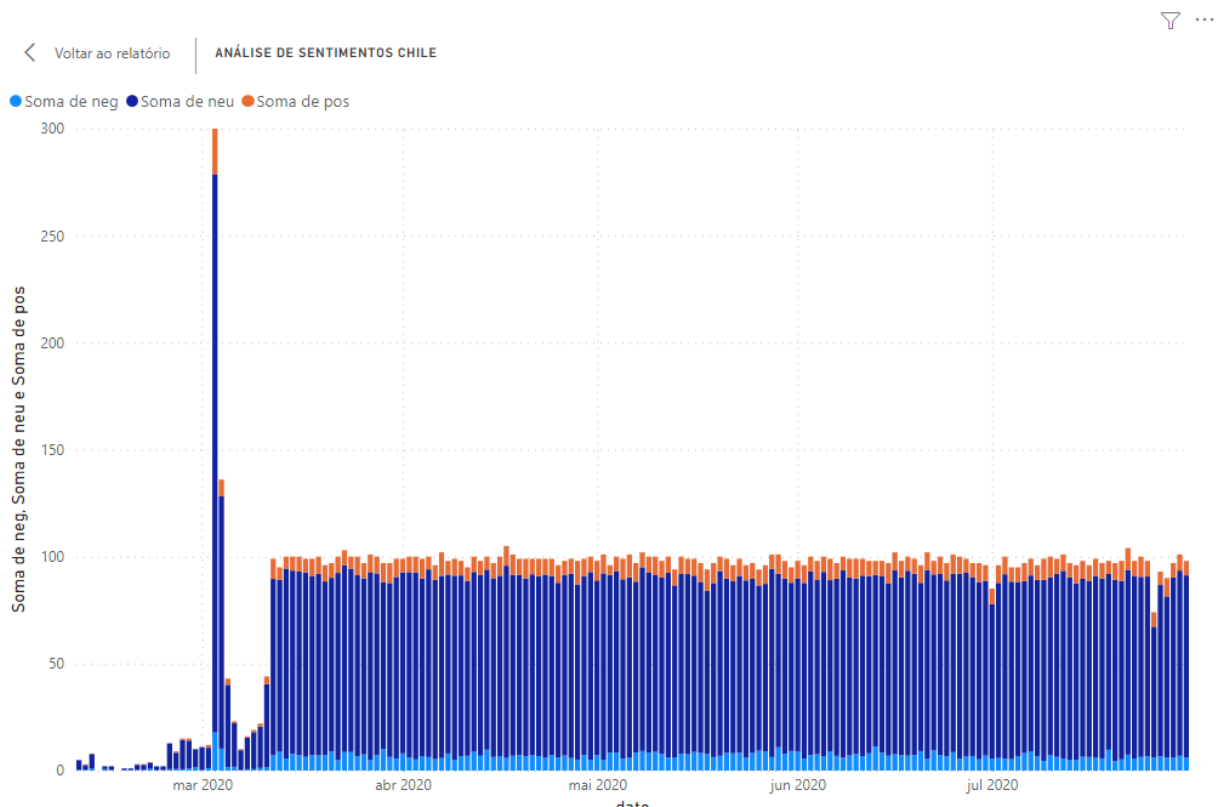


Figura 5 - Análise de sentimentos Chile

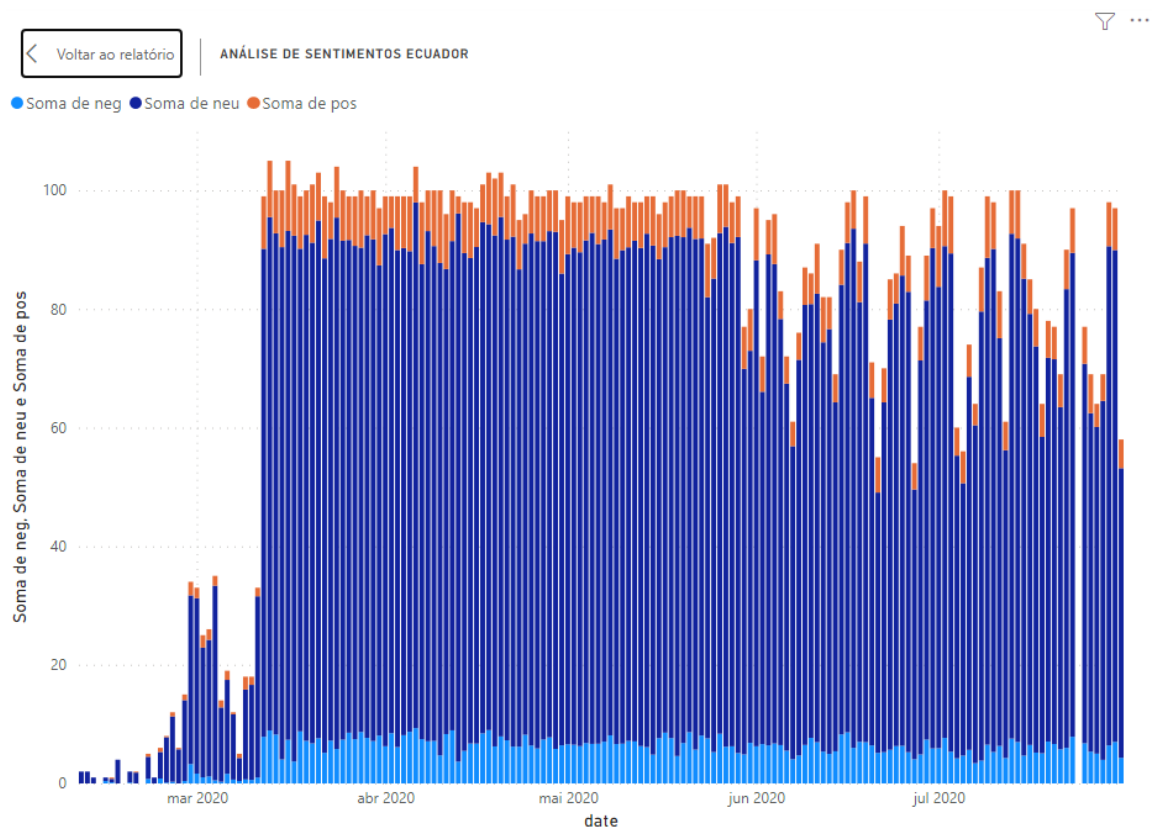


Figura 6 - Análise de sentimentos Equador

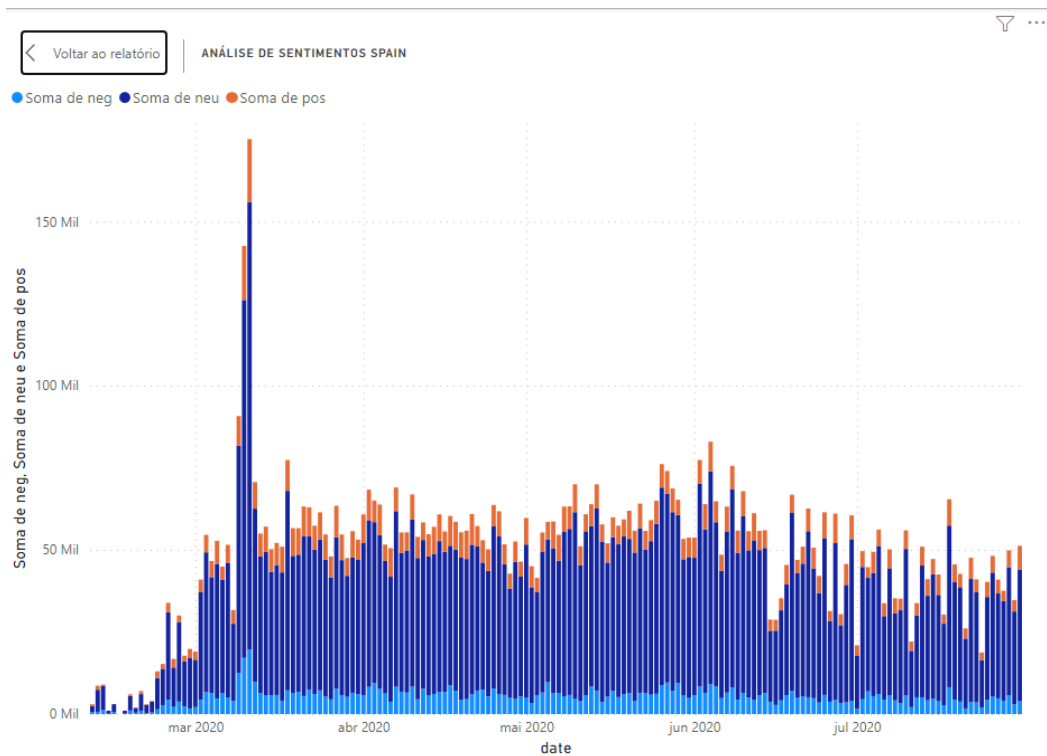


Figura 7 - Análise de sentimentos Espanha

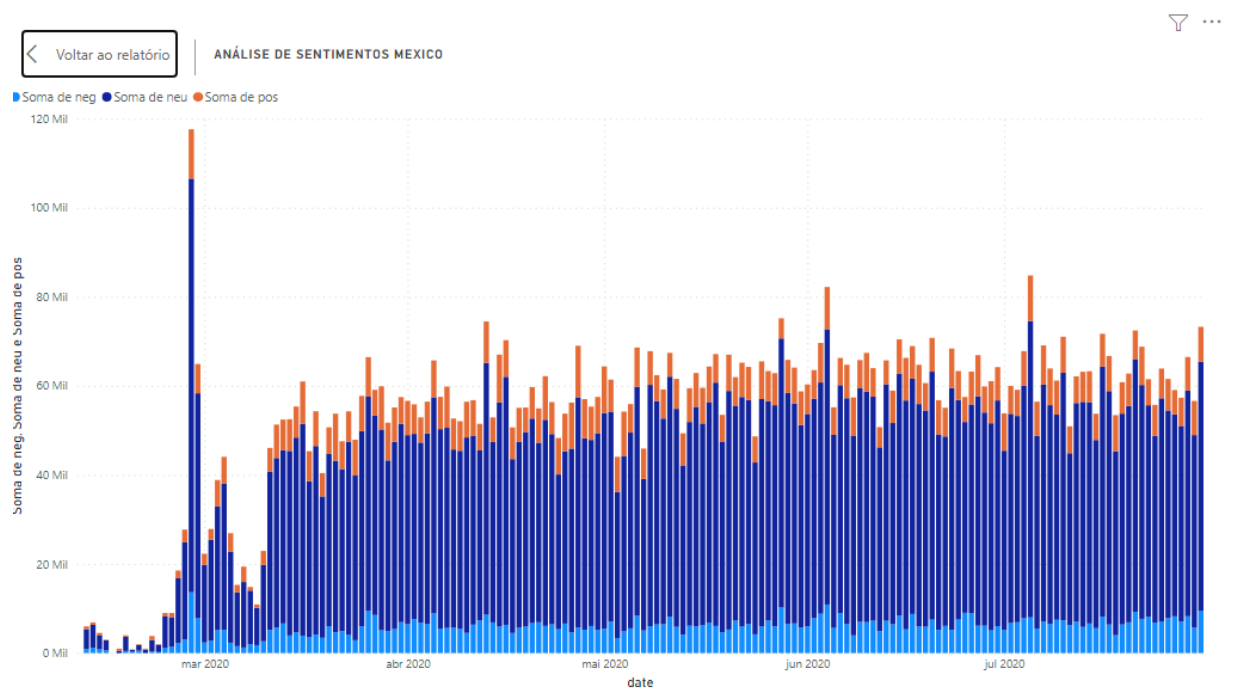


Figura 8 - Análise de sentimentos México

3.2. Previsão do número de casos de COVID modelo ARIMA

3.2.1. Metodologia

O modelo ARIMA (Autorregressivo Integrado de Médias Móveis) é a combinação dos modelos AR e MA com a adição de um termo de diferenciação. ARIMA é uma classe de modelos que “explica” uma determinada série temporal com base nos seus próprios valores passados. Os termos de um modelo ARIMA são p, d e q.

O termo “p” refere-se à componente AR do modelo, já mencionada anteriormente. O termo “q” refere-se à componente MA do modelo, também já mencionado. E o termo “d” é a ordem de diferenciação necessária para que a série temporal seja estacionária.

A prática mais comum para tornar uma série estacionária é diferenciá-la. Ou seja, subtrair o valor corrente do valor anterior. Para algumas séries é necessário realizar mais de uma diferenciação. Portanto, o valor de “d” é a quantidade mínima de diferenciações para que a série seja estacionária, caso ela já seja estacionária o valor de “d” é igual a zero.

No modelo ARIMA a série foi diferenciada pelo menos uma vez e sua formulação matemática é:

$$y(t) = \alpha + p_1 y_{t-1} + p_2 y_{t-2} + \dots + p_i y_{t-i} + \epsilon_t + q_1 \epsilon_{t-1} + q_2 \epsilon_{t-2} + \dots + q_i \epsilon_{t-i} + \varepsilon$$

Em termos literais seria:

Valor predito de y no instante t = Uma constante + combinação linear de lags de y + combinação linear de erros de previsão com lags + ruído branco.

3.2.2. Resultados

A Figura 9 mostra o comportamento do modelo para a previsão de novos casos, para a média de 7 dias, em cada um dos países de estudo. No geral observa-se que o modelo ARIMA consegue representar bem a tendência de novos casos, tanto positiva quanto negativa.

Para as séries da Argentina e do México, segundo a Figura 9, o modelo tende a subestimar os valores observados. Para Chile e Equador as previsões geradas ficam, em média, abaixo dos valores observados. Já para a Espanha não se apresenta um viés claro na Figura 9.

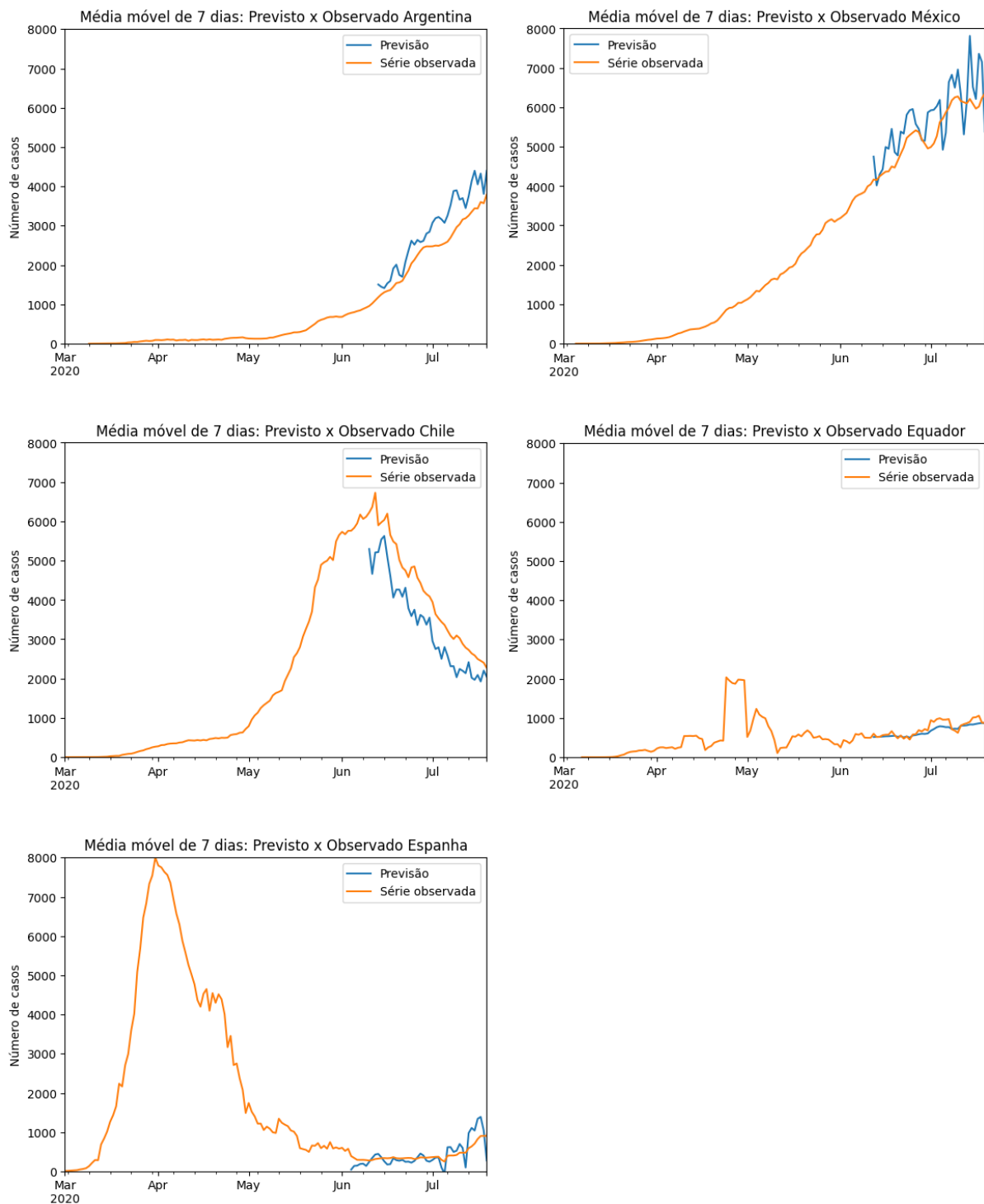


Figura 9 - Comparativo entre os valores da média móvel de 7 dias para casos previstos e observados.

Uma avaliação mais detalhada do modelo é apresentada na Tabela 1. Na Tabela 1 são apresentadas algumas das principais métricas de avaliação do modelo por horizonte de previsão. Cada horizonte refere-se a quantos dias à frente foram previstos. Por exemplo: h1 significa que foram previstos casos para 1 dia a frente, ou

seja, se o modelo fosse rodado hoje esta seria a previsão para o número de novos casos amanhã; o horizonte h_2 significa dois dias a frente, novamente, se o modelo fosse rodado hoje esta seria a previsão de novos casos para depois de amanhã e assim por diante.

Para as métricas de correlação e R^2 quanto mais próximos de 1 for seu valor melhor o desempenho do modelo. Para a métrica MAE quanto menor o seu valor melhor o modelo. A magnitude da métrica MAE é dependente do número de casos observados, portanto, países com números de casos observados maiores tendem a ter valores de MAE maiores. Desta forma esta métrica não deve ser considerada para avaliação do desempenho do modelo entre países. Já as métricas R^2 e Correlação, como assumem valores padronizados, permitem a comparação do desempenho do modelo entre países.

Assim, observa-se que a medida que se aumenta o horizonte de previsão o modelo tende a ter redução no seu desempenho. Isto ocorre devido ao aumento das incertezas associadas e já se é esperado que ocorra. Quanto mais longe estamos da data inicial maiores são as incertezas associadas e por isto, em geral, os modelos de previsão tendem a ter resultados piores. Por isto não se pode esperar que modelos sejam bons para realizar previsões para longos períodos de tempo.

Os piores desempenhos foram registrados para México e Equador. Na análise preliminar, para definição dos parâmetros dos modelos ARIMA para o Equador foi verificado que os casos registrados neste país, não têm forte dependência temporal. Isto pode justificar o baixo desempenho do modelo para este país. Para o caso do México o modelo apresenta uma sazonalidade, Figura 9, que não é registrada na série observada. Já o país de melhor desempenho, por horizonte de previsão, foi a Argentina. Em geral, o modelo utilizado apresenta melhor desempenho para a previsão da média de 7 dias.

Tabela 1 - Principais métricas de avaliação do modelo ARIMA.

País	horizonte	R ²	MAE	Correlação
Argentina	h1	0,78	315,97	0,89
Argentina	h2	0,74	387,90	0,87
Argentina	h3	0,62	452,51	0,80
Argentina	h4	0,65	462,36	0,82
Argentina	h5	0,67	477,85	0,83
Argentina	h6	0,70	459,59	0,86
Argentina	h7	0,70	466,88	0,85
Argentina	7 dias	0,84	278,59	0,92
Chile	h1	0,51	885,92	0,74
Chile	h2	0,53	863,68	0,74
Chile	h3	0,53	809,36	0,75
Chile	h4	0,37	800,55	0,68
Chile	h5	0,71	497,92	0,86
Chile	h6	0,68	521,73	0,85
Chile	h7	0,58	572,12	0,79
Chile	7 dias	0,90	280,28	0,96
Equador	h1	-0,05	323,35	0,15
Equador	h2	-0,03	340,08	0,22
Equador	h3	-0,05	342,66	0,20
Equador	h4	-0,04	341,95	0,23
Equador	h5	-0,05	336,56	0,23
Equador	h6	-0,05	343,87	0,26
Equador	h7	-0,07	350,81	0,22
Equador	7 dias	-0,28	157,48	0,57
Espanha	h1	0,08	319,32	0,32
Espanha	h2	0,14	359,06	0,41
Espanha	h3	0,33	341,00	0,64
Espanha	h4	0,25	375,95	0,58
Espanha	h5	0,16	427,71	0,51
Espanha	h6	0,31	397,21	0,66
Espanha	h7	0,40	370,93	0,76
Espanha	7 dias	0,45	215,61	0,80
México	h1	-0,05	869,31	0,57
México	h2	-0,59	1.078,48	0,30
México	h3	-0,12	948,99	0,36
México	h4	-0,25	995,14	0,27
México	h5	-0,22	986,59	0,24
México	h6	0,32	769,87	0,57
México	h7	0,38	730,07	0,62
México	7 dias	-0,03	527,70	0,68

4. Dados Complementares

Para complementar o estudo foram coletados dados sobre o mercado de energia dos países em estudo e também informações sobre as datas de lockdown em cada país.

4.1. *Mercado de energia*

A Figura 10 mostra a demanda mensal de energia em TWh/mês entre 2019 e 2021 para os países em estudo. Optou-se em colocar cada país em um gráfico separado pois a diferença de magnitude entre eles não permitiria visualizar mudanças nas séries, dos países de menor demanda, caso fossem colocados todos no mesmo gráfico e mesma escala.

No início da pandemia muito se falou que as medidas de lockdown iriam, no quesito economia, trazer mais malefícios que benefícios. A alegação era de que parar as atividades econômicas iria afetar a economia dos países a ponto que seria difícil se recuperar depois. Em eventos que causem tantos impactos na sociedade o mercado de energia acaba sendo um dos primeiros a ser afetado e também um dos primeiros a se recuperar. Então a variação nas séries de demanda de energia elétrica acaba sendo um ótimo termômetro de como está a atividade econômica de um país. Quanto maior atividade econômica maior a demanda por eletricidade. Vale salientar que, embora a maior quantidade de usuários de energia de um país esteja no setor residencial são os setores industrial e comércio que geram a maior demanda de energia elétrica de um país.

Em todas as séries da Figura 10 observa-se que houve diminuição da demanda de energia no ano de 2020, inclusive aqueles que não adotaram medidas de lockdown. Porém, as mudanças ficam mais evidentes naqueles que adotaram medidas mais restritivas como Argentina e Espanha. Chile foi um país que não adotou medidas de lockdown e observa-se que, praticamente, durante todo o ano de 2020 a demanda de eletricidade foi a menor da série analisada. Isto pode ter sido efeito da pandemia no país. Argentina e Espanha, que adotaram as medidas mais restritivas, tiveram as maiores diminuições na demanda de energia elétrica mas também foram os países, nos quais, a demanda de energia elétrica voltou a patamares “normais” com maior velocidade.

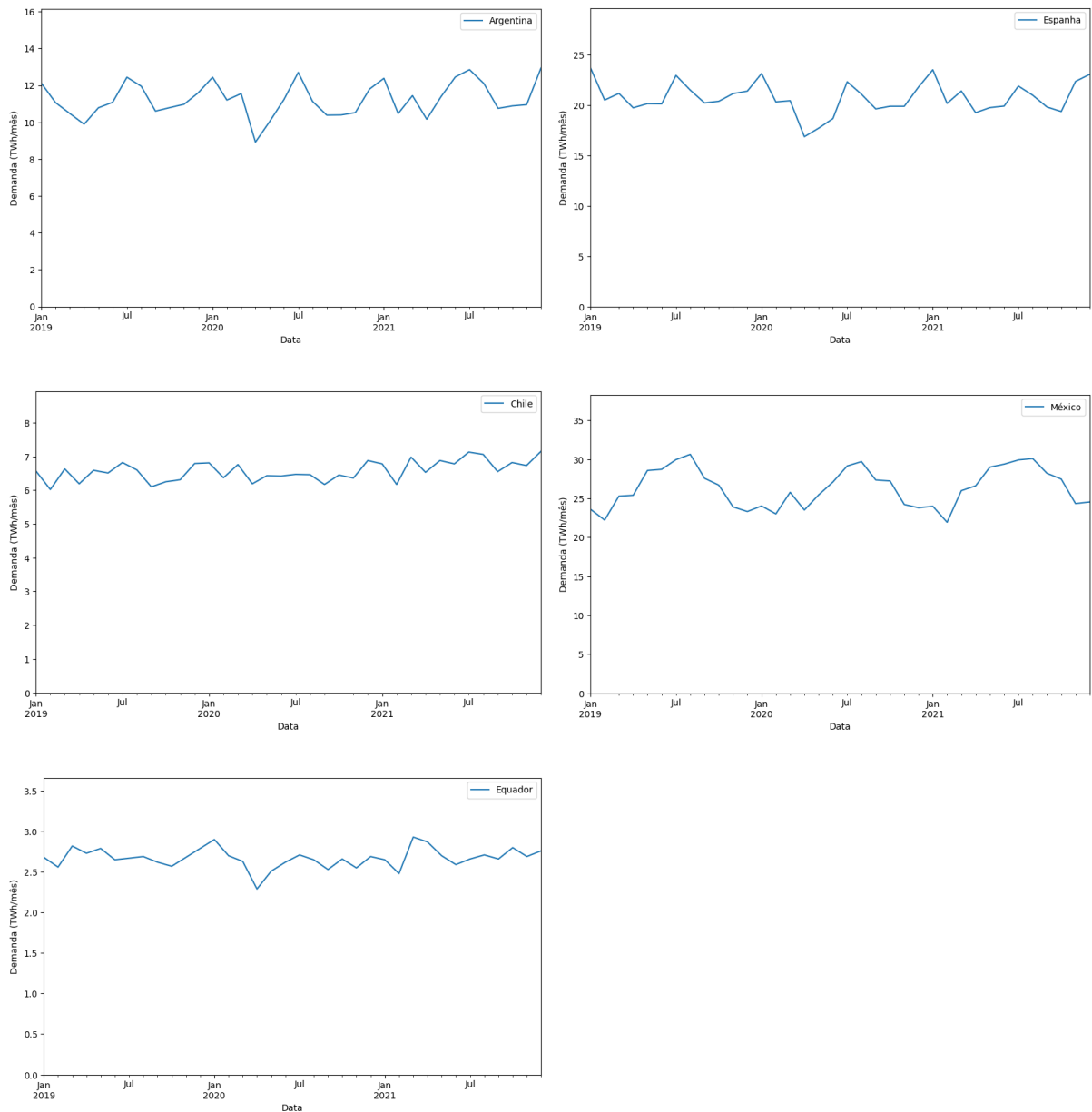


Figura 10 - Demanda mensal de energia.

4.2. *Datas de lockdown*

As datas de ocorrência de lockdown foram consideradas importantes pois trata-se de uma das primeiras medidas adotadas pelos países na tentativa de conter o avanço da COVID. Cada país adotou esta medida em períodos diferentes, alguns mais de uma vez, e com grau de controle sobre o deslocamento da população diferente. A ideia do lockdown é que com restrição do fluxo de pessoas também haja diminuição da dispersão do vírus e consequentemente da taxa de contágio.

A informação das datas de início e fim dos lockdowns foram obtidas do site da Wikipedia pelo seguinte link: (https://en.wikipedia.org/wiki/COVID-19_lockdowns). Na Tabela 2 estão registrados apenas os lockdowns a nível nacional de cada país nos primeiros meses da pandemia.

Tabela 2 - Dados dos lockdowns de cada país.

País	Data de Início	Data de Fim	Total de Dias
Argentina	19/03/2020	10/05/2020	52
Equador	16/03/2020	31/03/2020	15
México	23/03/2020	01/06/2020	70
Espanha	14/03/2020	09/05/2020	56

O Chile não adotou medidas rígidas de isolamento, a nível nacional, e ao longo do período algumas cidades em específico decretaram lockdown.

5. Dashboard Funcional

Após os ajustes realizados na pasta do medallion, refeito todo o dashboad em um arquivo novo, importando os dados diretamente do drive para ficar atualizado.

Abaixo planilhas importadas:

```

> [icon] avaliacao
> [icon] classificacao_hapiness
> [icon] country_wize_latest
> [icon] covid_19_clean_complete
> [icon] data
> [icon] demanda
> [icon] ember_electricity_mont...
> [icon] full_grouped_filtro
> [icon] lockdown
> [icon] nota_hapiness
> [icon] paises
> [icon] twitter_AR
> [icon] twitter_CL
> [icon] twitter_EC
> [icon] twitter_sentimental_ana...
> [icon] twitter_sentimental_ana...
> [icon] worldometer_filtro

```

- Avaliação: Prévia das métricas do modelo preditivo ARIMA
- Classificacao_hapiness: Classificação do índice de felicidade dos países selecionados;
- country_wize_latest: valores acumulados do período dos países.
- covid_19_clean-complete: resumo das informações covid e dos países.
- data: colocado as datas separados pois estavam em conflito,
- demanda: os dados da demanda de energia do período de covid.
- ember_eletricity_mont: informações de energia elétrica no período e dos países solicitados;
- full_grouped_filtro: informações de Covid dos países selecionados, em grupo.
- lockdown: Datas em que os países pararam para diminuir a transmissão do vírus;
- nota_hapiness: nota dada para os países no índice de felicidade;
- twitter_AR: análise dos sentimentos Argentina
- twitter_CL: análise dos sentimentos Chile
- twitter_EC: análise dos sentimentos Ecuador;
- twitter_ES: análise dos sentimentos Espanha;
- Twitte_MX: análise dos sentimentos México;
- worldometer_filtro: informações da covid e dos países.

As planilhas relacionadas:

Calculamos as taxas:

Taxa Positividade	Mortalidade	Letalidade
25,52%	0,04%	12,23%
2,08%	0,05%	2,53%
0,50%	0,06%	7,86%
2,17%	0,01%	1,84%
3,15%	0,03%	6,79%

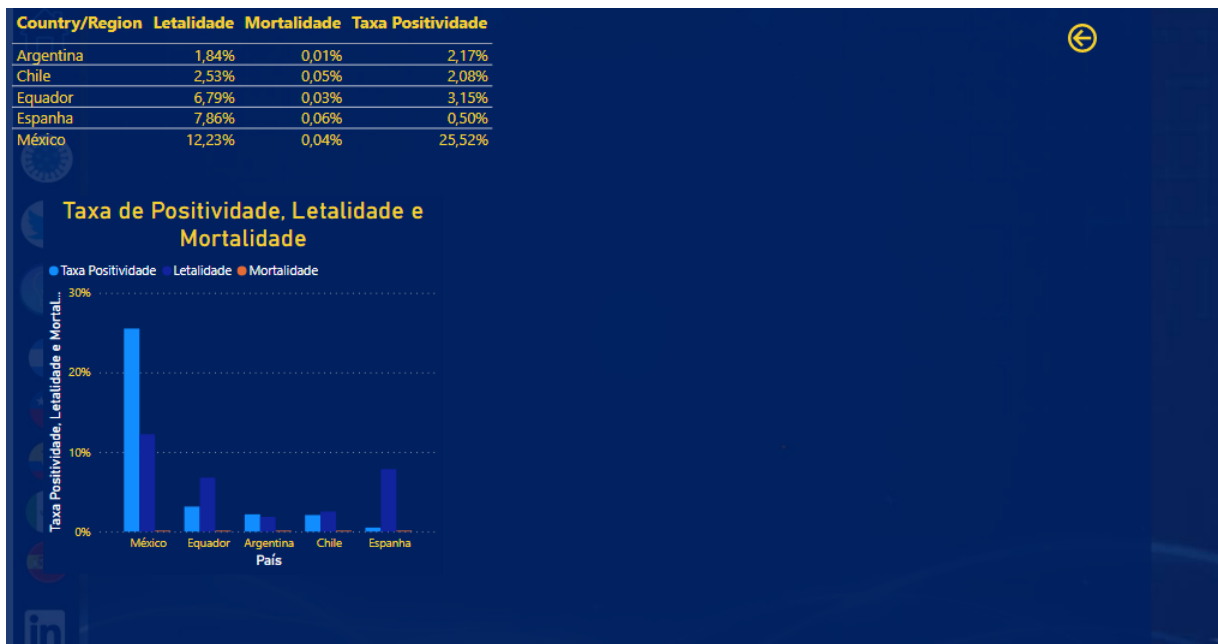
Dashboard - Covid:



- Cartão com a quantidade de casos confirmados dos cinco países selecionados;
 - 1 milhão de casos confirmados nos cinco países.
- Cartão com a totalidade de mortes dos países selecionados;
 - 95 mil no total de mortes nos cinco países no período de janeiro a julho 2020.
- Cartão com a quantidade de casos recuperados no período de janeiro a julho de 2020;
 - 899 mil pessoas recuperadas.
- Gráfico de testagem positivo, das pessoas que testaram para covid quantas realmente estavam positivadas;
 - 1,35% são de casos confirmados dentre todos os testados
- Gráfico de casos confirmados, no geral dos cinco países;
- Gráfico de casos de mortes no cálculo geral;
- Gráfico de casos confirmados por 1 milhão de habitantes;
- Gráfico de casos de mortes por 1 milhão de habitantes;

Analisando os gráficos, no contexto geral o México é o país com mais casos e mais mortes, entretanto quando observamos pela população a Argentina está em um estado mais grave. Pois a população do México é muito maior que a da Argentina.

No ícone de “*Clique para mais Informações*” é um indicador selecionado, abrindo como uma página separada.

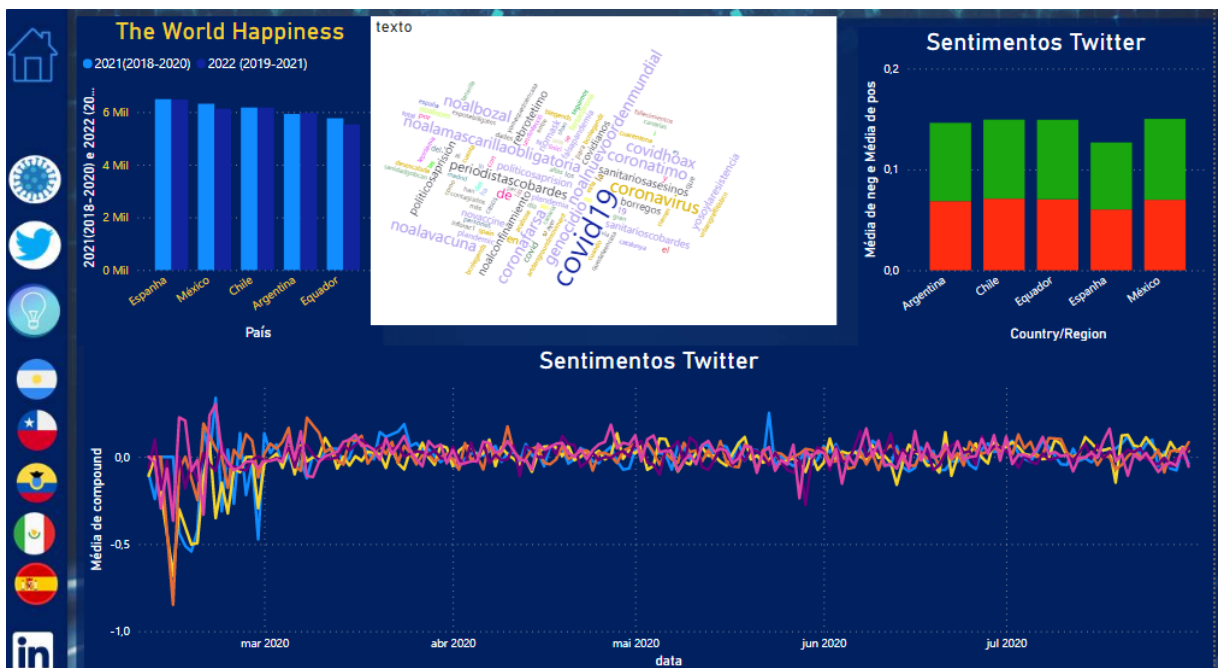


Nesta página contém as taxas de letalidade, mortalidade e positividade, e o gráfico representando isso em casa País.



No segundo slide, temos:

- Gráfico de evolução da doença, num contexto acumulativo;
- Média de novos Casos, esse é a Média ponderada dos dias, não é um acumulado;
- Médias de novas mortes, média ponderada dos dias, não é acumulado;



Sobre os dados coletados do twitter do período relacionado, e após o tratamento dos sentimentos e análises coletadas, podemos observar que os sentimentos do Twitter houve certa confusão no começo do período quando a organização mundial da saúde decretou que o Covid-19 se tornou uma pandemia.

O índice relacionado no gráfico, é o compound que vai de 1 a -1.

Também representamos o índice de felicidade mundial, para avaliar se houve uma baixa do índice no período pandêmico.

E relacionamos uma nuvem de palavras do twitter, está representado pelo da Espanha, pois foi o que teve mais riqueza em informações.

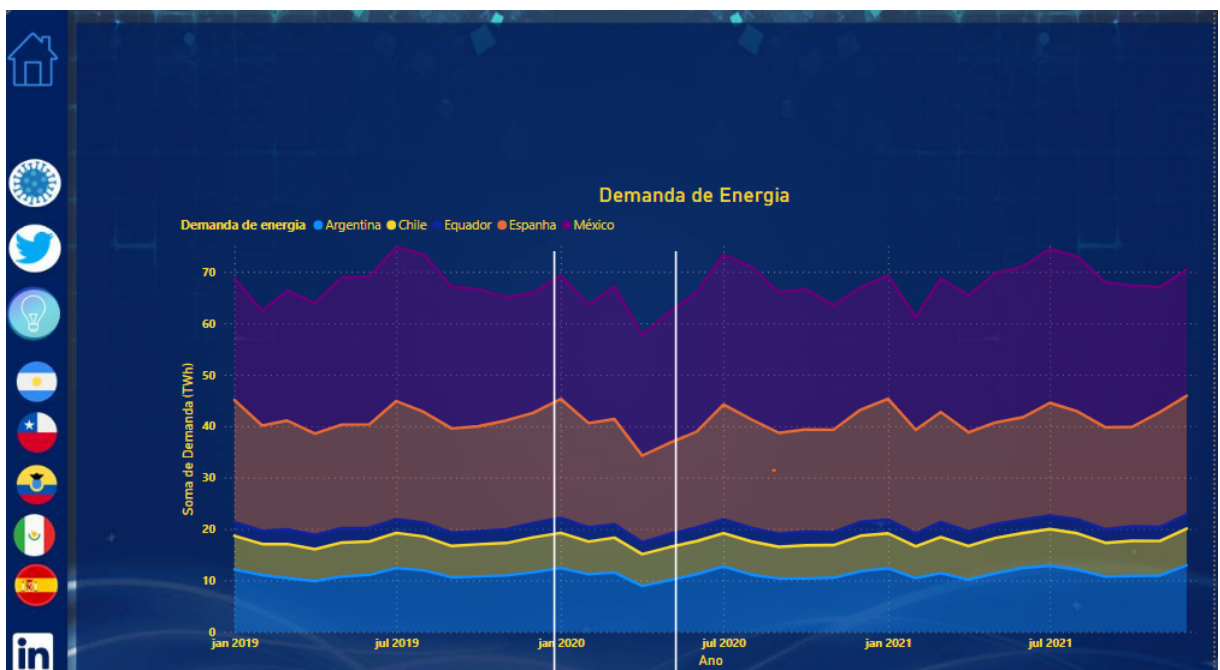
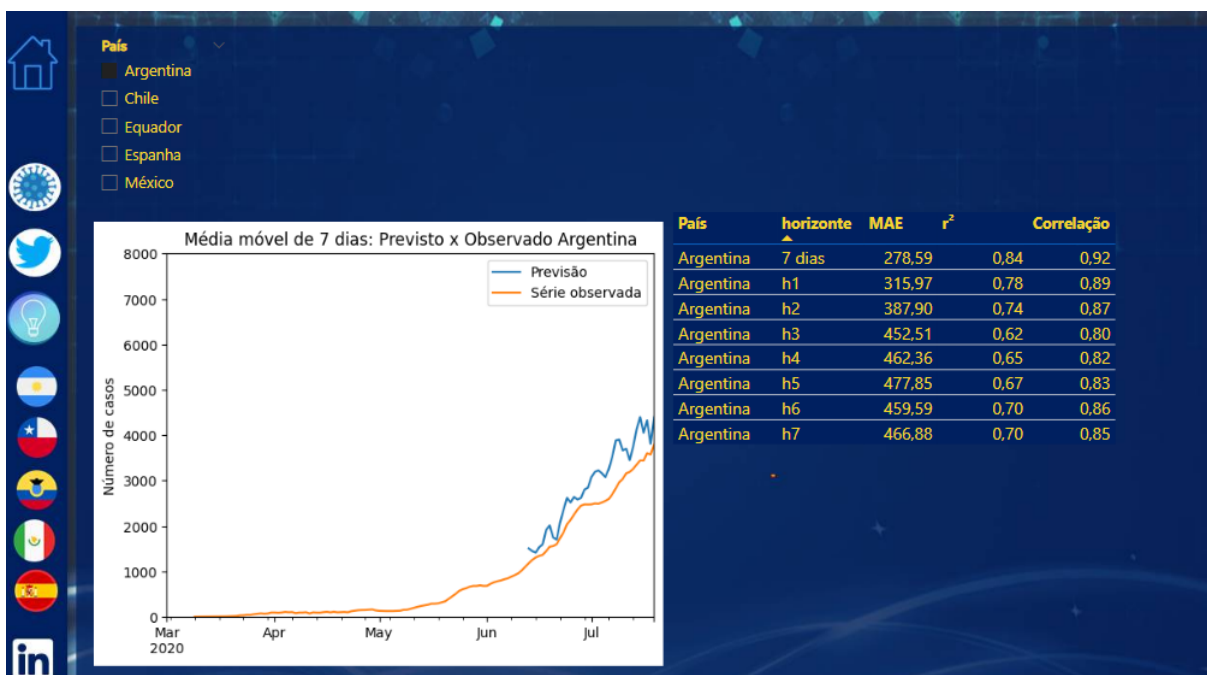


Gráfico da demanda de energia, para verificarmos se houve redução de consumo de energia no período de lockdown, pois nesse período quando está relacionado a parada econômica é o primeiro a sentir os impactos.



Prévia do gráfico do modelo de predição da média móvel de sete dias.

Tabela contendo as métricas obtidas numa observação de sete dias para frente de cada País.

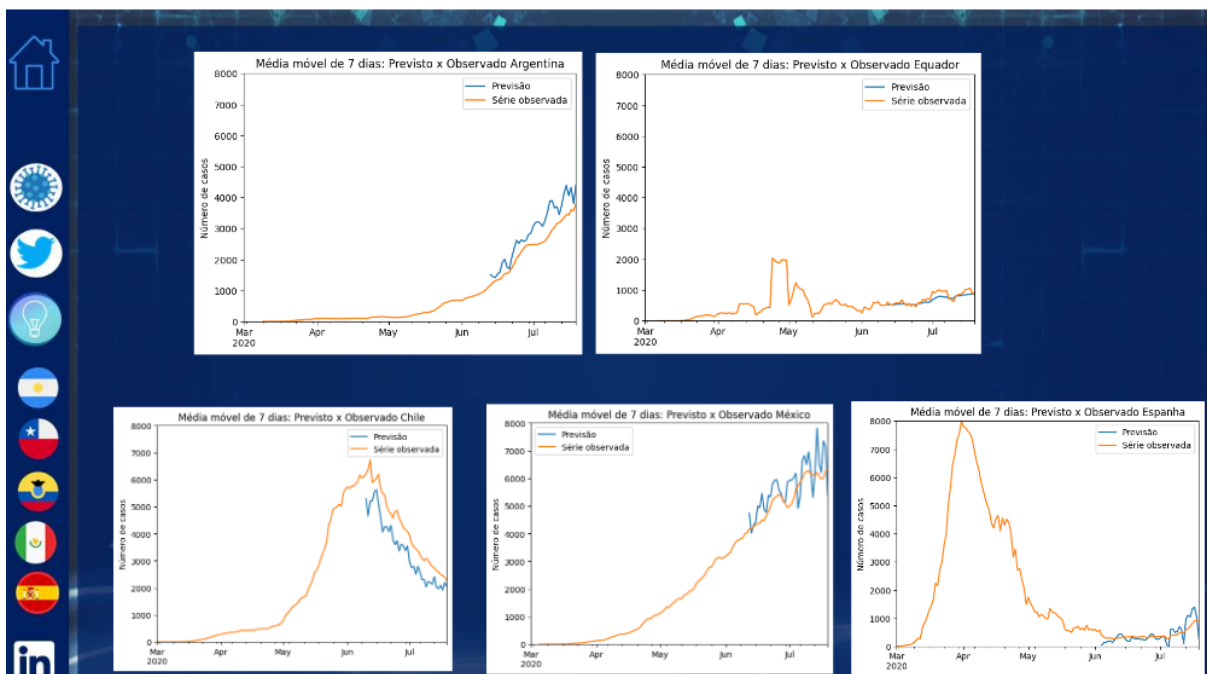


Gráfico dos modelos preditivos de todos os Países.

Dashboard ainda em desenvolvimento.

6. Tarefas para o próximo sprint.

1. Avanço no dashboard:
 - a. Pretende-se evoluir na interação do dashboard adicionando funcionalidades que o deixe mais interativo.
2. Avanço na arquitetura de dados:
 - a. Pretende-se criar um data warehouse na pasta gold para armazenamento e disponibilização dos dados para o dashboard.