



ANALISIS EPIDEMIOLOGICO MUNDIAL

VIRUS SARS- COV 2

2022

Mariano Sellan

CODERHOUSE



En este trabajo se va a realizar un análisis de la situación epidemiológica de cada región del mundo para así poder identificar cuáles son las regiones más afectadas por el brote de Covid 19.

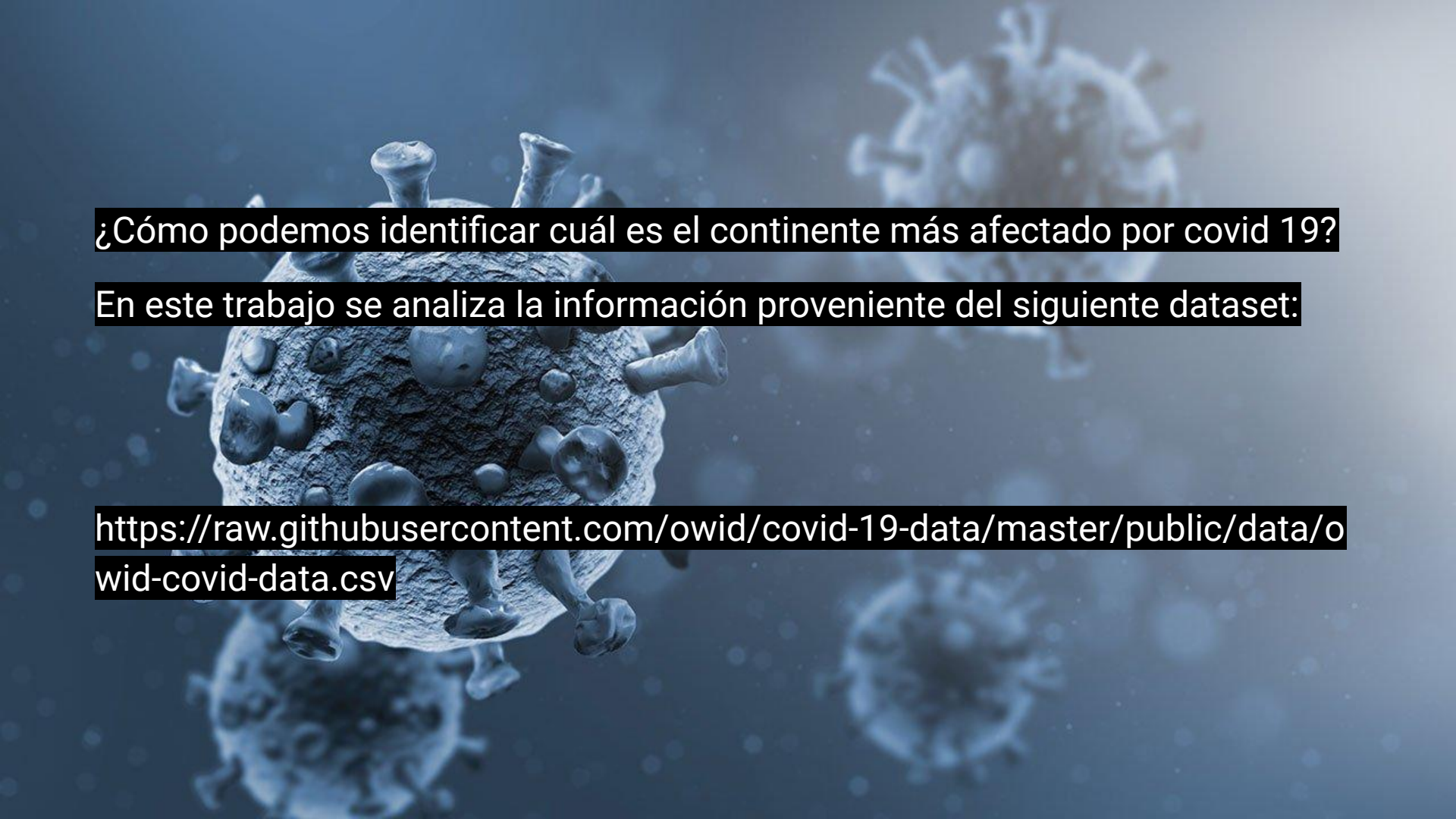
****Es importante destacar que los gráficos expuestos en este trabajo corresponden a una foto de la situación epidemiológica del mundo durante el periodo de agosto del 2022***



ABSTRACT

La ONU en conjunto con el Banco Mundial nos contrató para definir cómo será la distribución del primer envío de vacunas a nivel mundial, con el fin de brindar ayuda a los continentes más afectados por la pandemia.

Es por esto que se realizan diferentes gráficos con el fin de poder analizar la situación epidemiológica de distintas regiones/continentes del mundo. La principal hipótesis a resolver es:



¿Cómo podemos identificar cuál es el continente más afectado por covid 19?

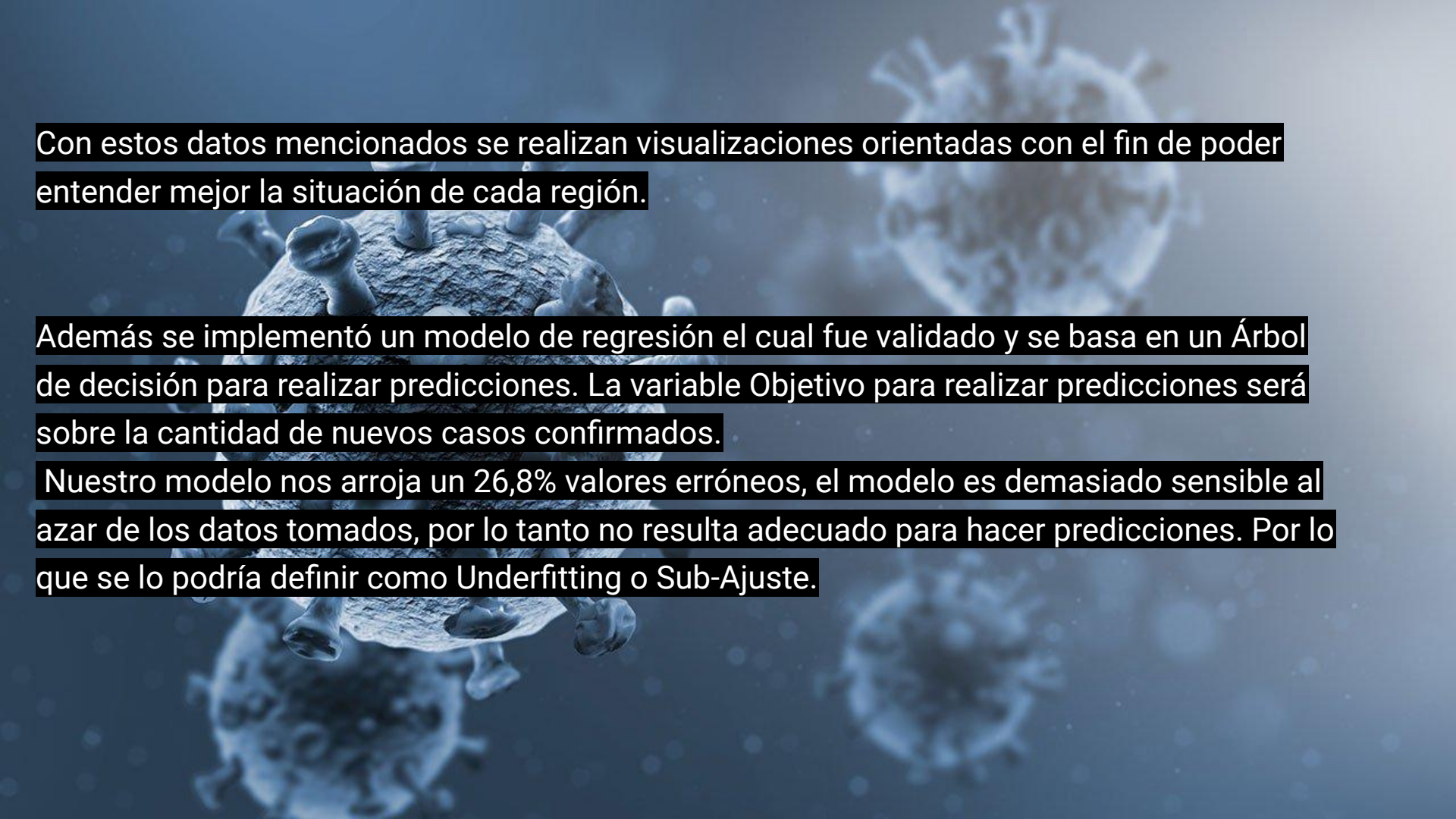
En este trabajo se analiza la información proveniente del siguiente dataset:

<https://raw.githubusercontent.com/owid/covid-19-data/master/public/data/owid-covid-data.csv>

Columnas de Interés:



Continent	Continente	Descripción
positive_rate	tasa de positividad	El porcentaje de test positivos sobre el total que se realizan
location	pais	
date	fecha	
total_cases	casos totales	casos acumulados
new_cases	nuevos casos	casos reportados en las ultimas 24 hs
people_vaccinated	personas vacunadas	
total_deaths	muerres totales	muerres acumuladas
new_deaths	nuevas muerres	muerres reportadas recientes (24hs)
median_age	edad media	

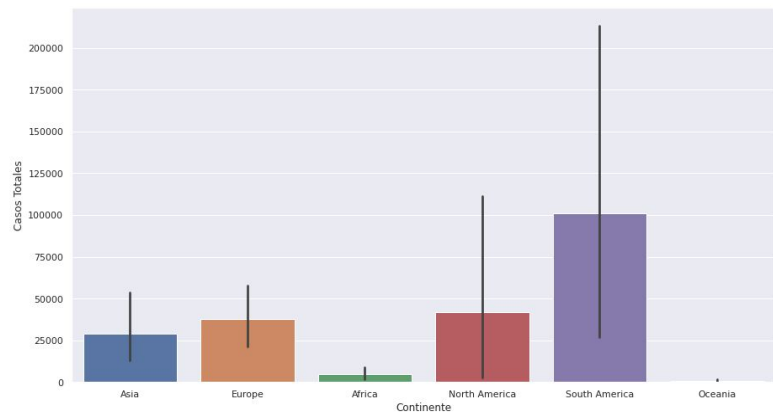
The background of the slide features a blue-tinted, microscopic image of several virus particles. These particles are spherical with a textured surface and numerous small, protruding spikes or receptors. They are scattered across the frame, with some appearing more prominent than others, creating a sense of depth and scientific focus.

Con estos datos mencionados se realizan visualizaciones orientadas con el fin de poder entender mejor la situación de cada región.

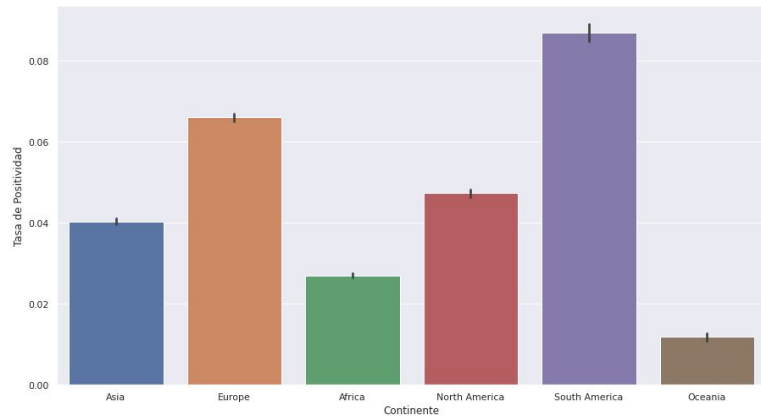
Además se implementó un modelo de regresión el cual fue validado y se basa en un Árbol de decisión para realizar predicciones. La variable Objetivo para realizar predicciones será sobre la cantidad de nuevos casos confirmados.

Nuestro modelo nos arroja un 26,8% valores erróneos, el modelo es demasiado sensible al azar de los datos tomados, por lo tanto no resulta adecuado para hacer predicciones. Por lo que se lo podría definir como Underfitting o Sub-Ajuste.

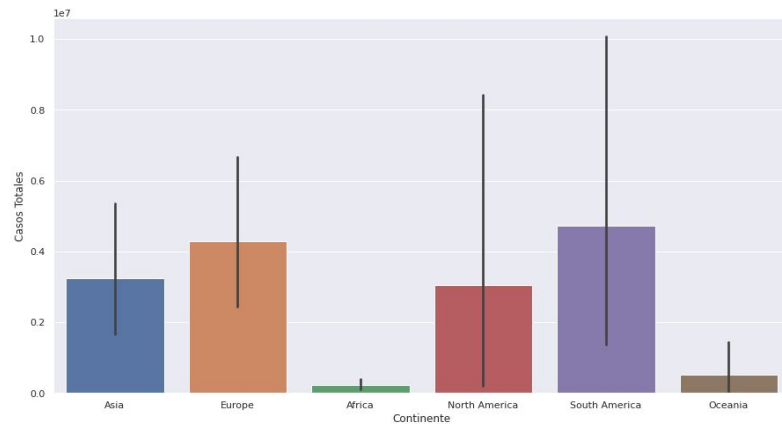
Muertes Reportadas



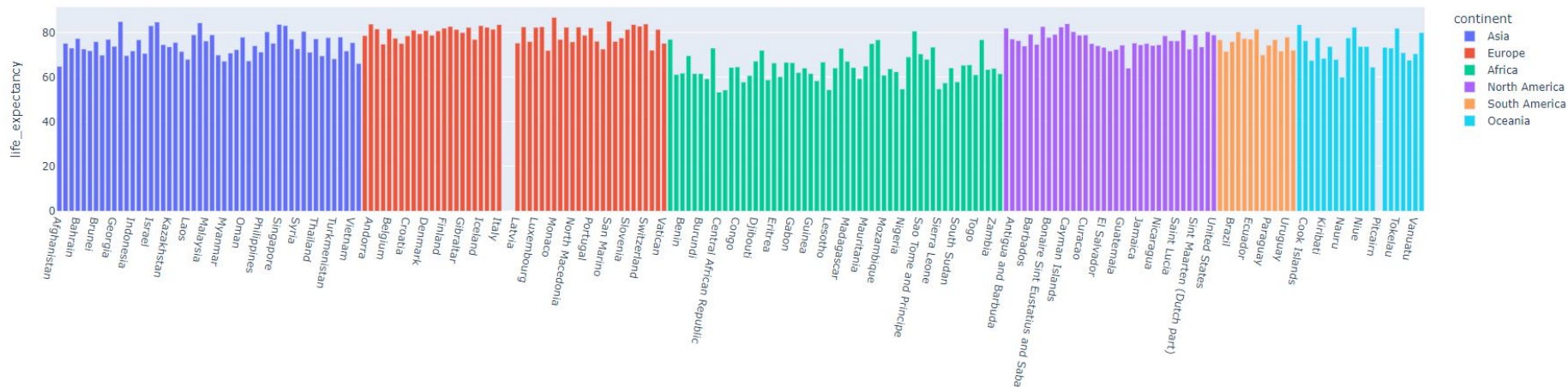
Tasa de Positividad



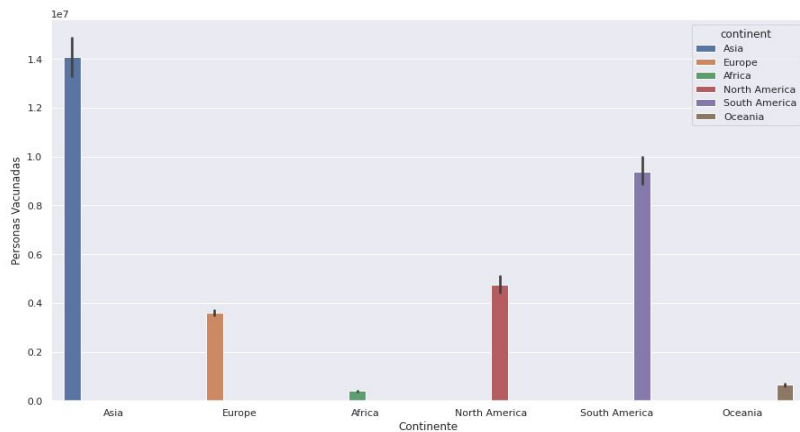
Casos totales reportados



Expectativa de Vida



Población Vacunada



Creación de Modelo de Machine Learning y su entrenamiento



#VALIDATION MODEL

```
##covid_file_path = "https://raw.githubusercontent.com/owid/covid-19-data/master/public/data/owid-covid-data.csv"
```

```
##df_real = pd.read_csv(covid_file_path)
```

```
y = df_real["new_cases"]
```

```
feature_columns = [ 'total_deaths', 'total_cases', 'new_deaths' ]
```

```
X = df_real[feature_columns]
```

```
X.fillna(0,inplace=True)
```

```
X
```

```
#Specify Model
```

```
covid_model = DecisionTreeRegressor()
```

```
#Fit Model
```

```
covid_model.fit(X,y)
```

```
print("First in-sample prediction:", covid_model.predict(X.head()))
```

```
print("Actual target values for those homes:", y.head().tolist())
```

ENTRENAMIENTO DE NUESTRO MODELO

```
[ ] # Import the train_test_split function
    from sklearn.model_selection import train_test_split

    train_X, test_X, train_y, test_y= train_test_split(X,y,test_size= 0.2)
```

```
▶ # Specify the model

    covid_model = DecisionTreeRegressor(random_state=1)

    # Fit iowa_model with the training data.
    covid_model.fit(train_X, train_y)
```

```
↳ DecisionTreeRegressor(random_state=1)
```

```
[ ] # Predict with all validation observations
    test_predictions = covid_model.predict(test_X)
```

```
[ ] # print the top few validation predictions
    print(test_predictions[:5])
    # print the top few actual prices from validation data
    print(test_y.values[0:5])
```

```
[ 884.    57.    0.  305. 10424.]
[ 679.   340.    0.   522.  9003.]
```

```
from sklearn.metrics import mean_absolute_error

test_mae = mean_absolute_error(test_y, test_predictions)

print(test_mae)
```

```
5634.255319148936
```

```
[ ] test_mae/y.mean() * 100

26.888080770522365
```

```
[ ]
```

Error absoluto medio (MAE)

Nuestro pronostico esta errado en un 26,8%. Lo que dice bastante sobre nuestro modelo, deberiamos poder ajustarlo para que su prediccion sea mas efectiva.

La variable Objetivo para realizar predicciones sera sobre la cantidad de nuevos casos confirmados.

El modelo es demasiado sensible al azar de los datos tomados, por lo tanto no resulta adecuado para hacer predicciones. Por lo que se lo podria definir como Underfitting o Sub-Ajuste.

- Se intento agregar mas columnas para realizar la prediccion, lo que arrojó valores aun mas grande de MAE, por lo que se descarto esa opcion
- Como opcion se podria probar algun otro modelo para ver si el resultado es similar.

Interpretación de resultados y conclusiones

- Como podemos deducir en base a nuestros insights la mayor cantidad de casos reportados se encuentra en la región de América del Sur
- Además en América del sur tenemos el número más alto de decesos reportados por covid.
- También podríamos afirmar que Asia viene vacunando a su población de una manera más rápida que el resto
- Podemos ver como Brasil (ubicado en el continente más afectado) tiene un ritmo de vacunación mucho más alto que los demás países de la región.
- Otra valoración interesante para hacer sobre los insights obtenidos es que la región que más testeos hizo fue América del Norte, pero América del Sur tiene un número inferior considerablemente lo que nos hace pensar que si la tasa de positividad es mayor claro está que si tuviéramos un mayor número de testeos la cantidad de casos totales aumentaron notablemente y también así las muertes reportadas.

CONCLUSIÓN

Analizando los gráficos y la información obtenida de los Insight podemos ver como América del Sur tiene la tasa más alta de positividad y sumado a esto, evaluamos casos puntuales de países de esta región donde la cantidad de muertes confirmadas es muy elevada en comparación con otras regiones.

Es por esto que se determina asignar un mayor porcentaje de vacunas a las zonas más afectadas

Asignaciones:

América del Sur = 22%

América del Norte = 16%

Oceanía = 12%

Asia = 15%

Europa = 16%

Africa = 19%

