

Predicción de inundaciones según datos de precipitaciones en las zonas Norte, Centro y Sur de Chile Continental

Estudiante: Carlos Hernández
Datascience Bootcamp - Coding Dojo

Introducción

Chile Continental está dividido por tres zonas geográficas: Norte, Centro y Sur. Las cuales están definidas por latitudes y longitudes [1]. Para cada zona existe un rango de milímetros de agua caída por concepto de lluvia percibidas en cuencas y sub cuencas [2].

Este proyecto se centra en las mediciones de precipitaciones percibidas y medidas por distintas estaciones meteorológicas distribuidas en las zonas geográficas mencionadas. Además, elabora y aplica modelos de Machine Learning para predecir si existe probabilidad de inundación según dichas mediciones.

[1] [Datos geográficos de Chile](#)

[2] [Definición de cuenca](#)



Imagen 1: Cuencas y sub cuencas [2]



Antecedentes de los datos

El conjunto de datos de milímetros caídos de precipitaciones en las tres zonas geográficas de Chile Continental procede de Center For Climate and Resilience Research ([sitio web](#)). Los datos son de libre acceso y uso y contiene registros de los 365 días del año desde el año 1950 al año 2020 [3], percibidos por distintas estaciones ubicadas a lo largo de Chile Continental, principalmente en zonas altas como cerros y montañas en la precordillera y cordillera de Los Andes

Este conjunto de datos originalmente tiene 37 características (variables), pero se hizo una selección de aquellas más relevantes para este estudio. Las variables que se usarán como objeto de estudio para este proyecto se darán a conocer a continuación.

[3] Datos de precipitación, Center For Climate and Resilience Research. URL:
<https://www.cr2.cl/datos-de-precipitacion/>

Descripción de las características



Nombre característica	Descripción
altura	Altura, expresada en metros sobre el nivel del mar, donde está instalada la estación que percibió precipitaciones.
latitud	Proporciona la localización de un lugar en dirección Norte o Sur desde el ecuador. Expresada en números decimales negativos. Se usa negativo para definir que está al sur del Ecuador.
longitud	Proporciona la localización de un lugar en dirección Este u Oeste desde el meridiano de Greenwich. Expresada en números decimales negativos. Se usa negativo para definir que está al Oeste de Greenwich.
codigo_cuenca	Código único para definir al nombre de una cuenca. Numérico.
codigo_sub_cuenca	Código único para definir al nombre de una sub cuenca. Numérico.
fecha	Fecha en el formato año-mes-día en que fueron registradas las precipitaciones.
mediciones	Milímetros de agua caída según fecha. Numérico, expresado en milímetros.

Limpieza:

```
regex = r'^19\d{2}-\d{2}-\d{2}$'
df_filtered = df_precipitaciones.loc[:, ~df_precipitaciones.columns.str.match(regex)]
```

Acortar dataframe para registros sólo desde el año 2000 al 2020

```
df_precip_clean.drop(['codigo_estacion', 'institucion', 'fuente'], axis=1, inplace=True)
```

Eliminar características

```
[ ] df_precip_clean = df_precip_clean[df_precip_clean['mediciones'] > 0]
```

Eliminar registros donde no hubo lluvia o medición errónea

```
df_precip_clean['latitud'] = df_precip_clean['latitud'].astype(float)*-1
df_precip_clean['longitud'] = df_precip_clean['longitud'].astype(float)*-1
```

Transformar latitud y longitud para mejor manera

```
def obtener_estacion_chile(fecha):
    mes = fecha.month
    dia = fecha.day
    if (mes == 12 and dia >= 21) or (mes in [1, 2]) or (mes == 3 and dia <= 20):
        return 1 #verano
    elif (mes == 3 and dia >= 21) or (mes in [4, 5]) or (mes == 6 and dia <= 20):
        return 2 #otoño
    elif (mes == 6 and dia >= 21) or (mes in [7, 8]) or (mes == 9 and dia <= 20):
        return 3 #invierno
    else:
        return 4 #primavera
```

Crear característica para estación del año según fecha

```
def determinar_zona(lat, lon):
    if (lat >= 17.0 and lat <= 26.0):
        return 1 #norte
    elif (lat > 26.0 and lat <= 35.0):
        return 2 #centro
    else:
        return 3 #sur

df_precip_clean['zona'] = df_precip_clean.apply(lambda row: determinar_zona(row['latitud'], row['longitud']), axis=1)
df_precip_clean['zona'] = df_precip_clean['zona'].astype(int)
```

Crear característica de zona geográfica: 1: Norte, 2: Centro, 3: Sur

```
[ ] df_precip_clean['mediciones'].value_counts()
```

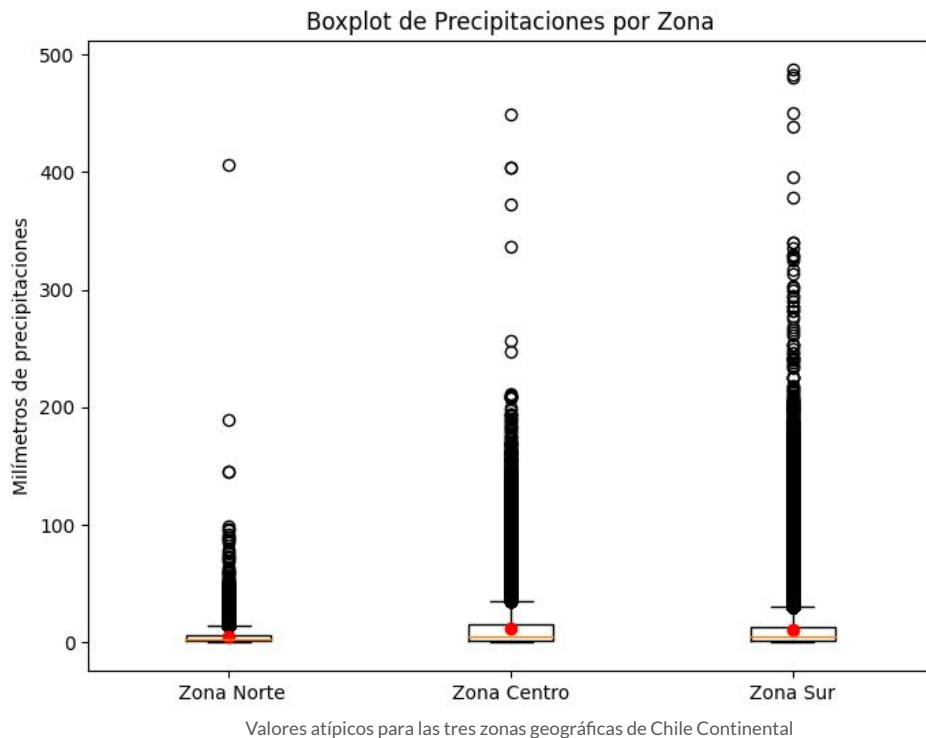
mediciones	
0.00	3229935
-9999.00	2400801
0.10	40912
1.00	32276
0.20	32017
...	...
208.70	1
154.10	1
137.70	1
37.35	1
7.16	1

Name: count, Length: 2661, dtype: int64

Identificar mediciones incorrectas

Limpieza

Identificar valores atípicos





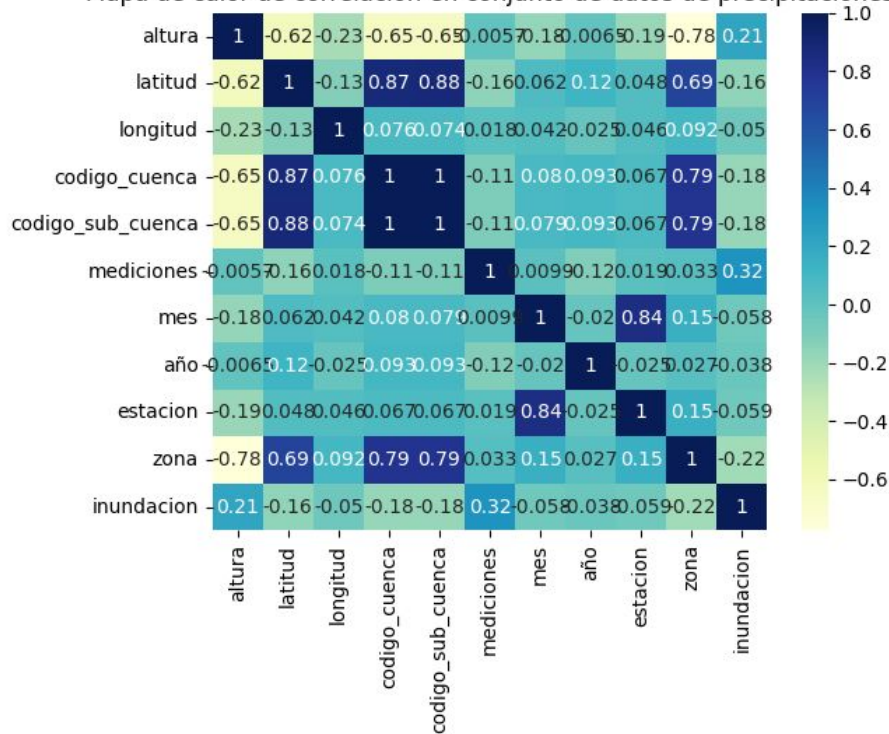
Definir variable objetivo (target)

Cada zona geográfica de Chile Continental tiene definido el rango de milímetros de precipitaciones esperados. Cuando supera ese límite, se considera inundación. Se considera inundación a subidas del nivel de agua dulce por sobre lo normal (límites).

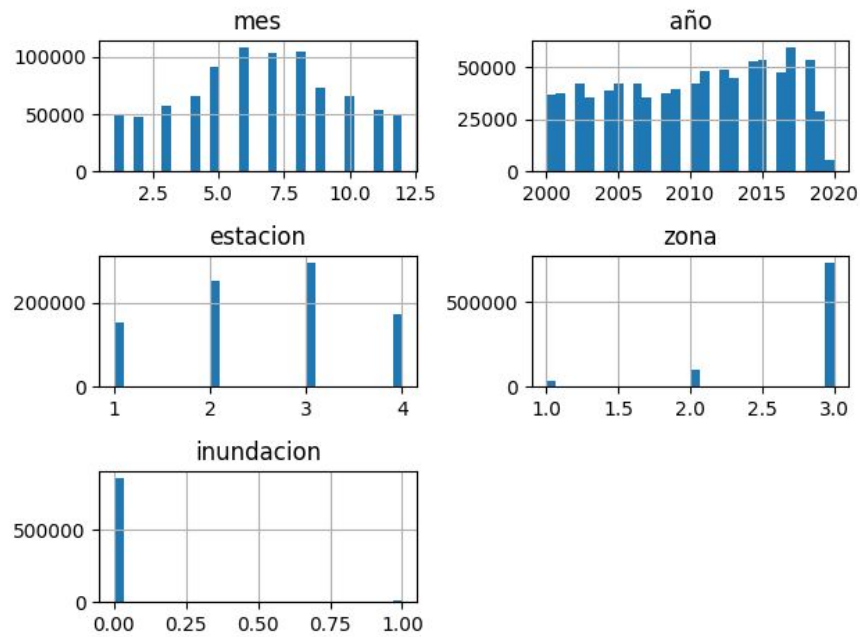
```
def definir_inundacion(mediciones, zona):  
    if mediciones >= 10 and mediciones <= 20 and zona == 1:  
        return 1  
    elif mediciones >= 50 and mediciones <= 100 and zona == 2:  
        return 1  
    elif mediciones >= 100 and mediciones <= 200 and zona == 3:  
        return 1  
    else:  
        return 0  
  
df_precip_clean['inundacion'] = df_precip_clean.apply(lambda row: definir_inundacion(row['mediciones'], row['zona']), axis=1)
```

Análisis Exploratorio de Datos (EDA)

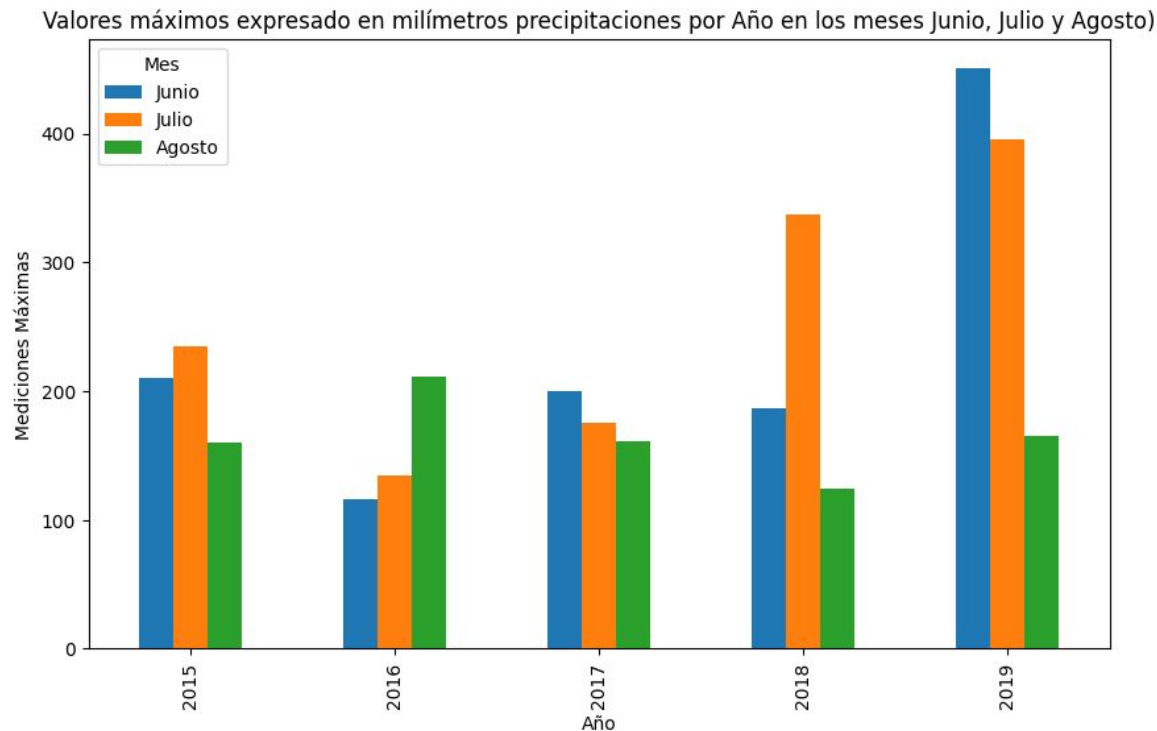
Mapa de calor de correlación en conjunto de datos de precipitaciones



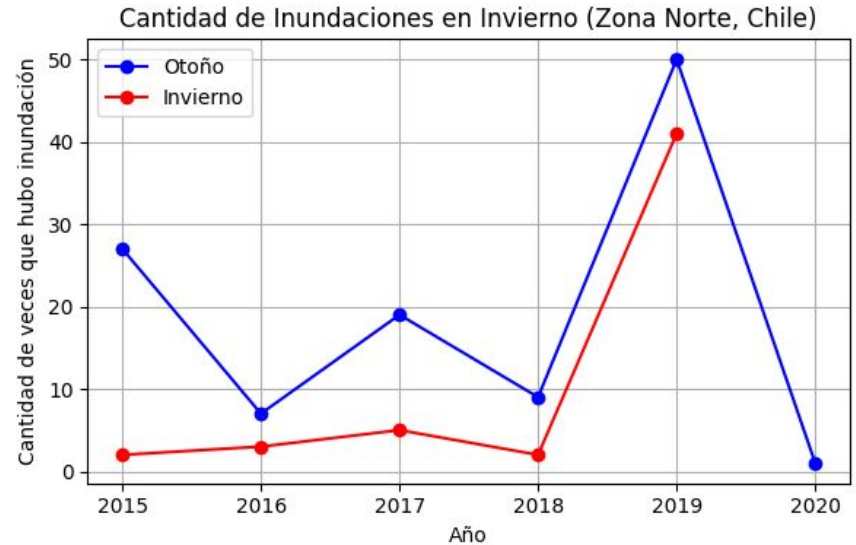
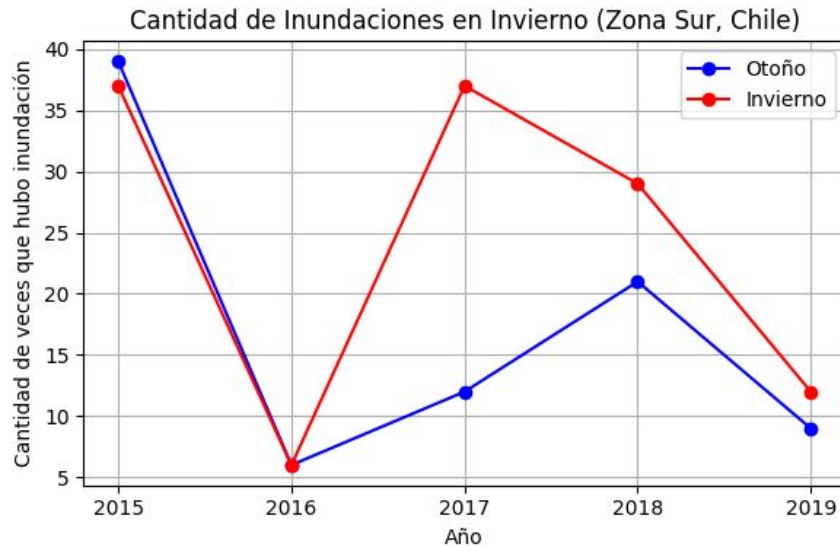
Histograma de todas las variables del conjunto de datos de precipitaciones



Análisis Exploratorio de Datos (EDA)



Análisis Exploratorio de Datos (EDA)





Desafíos y cómo fueron abordados

- Agrupar por mes y año para calcular máximos de precipitaciones registradas
- Explorar datos y descubrir qué meses son más lluviosos para las 3 zonas geográficas → gráficos, código: `groupby`, `value_counts()`