



Modelo predictivo de la demanda de bicicletas según horario en Capital Bikeshares DC

Presentado por: Andy Sanjur

Evaluated por: Juan Marcos Castillo, PhD



Motivación Principal



- **Comunidad deportiva**
- **Alternativa al transporte**
- **Proyección de volúmenes de alquiler**
- **Posibilidad de nuevos mercados**
- **Determinar condiciones que influyen en el público.**
- **Influencia de múltiples variables**



Extraído de: <https://fepaci.com.pa/wp-content/uploads/2023/03/DSC0801-1536x1024.jpg>

Conjunto de datos



- **Mostrar patrones**

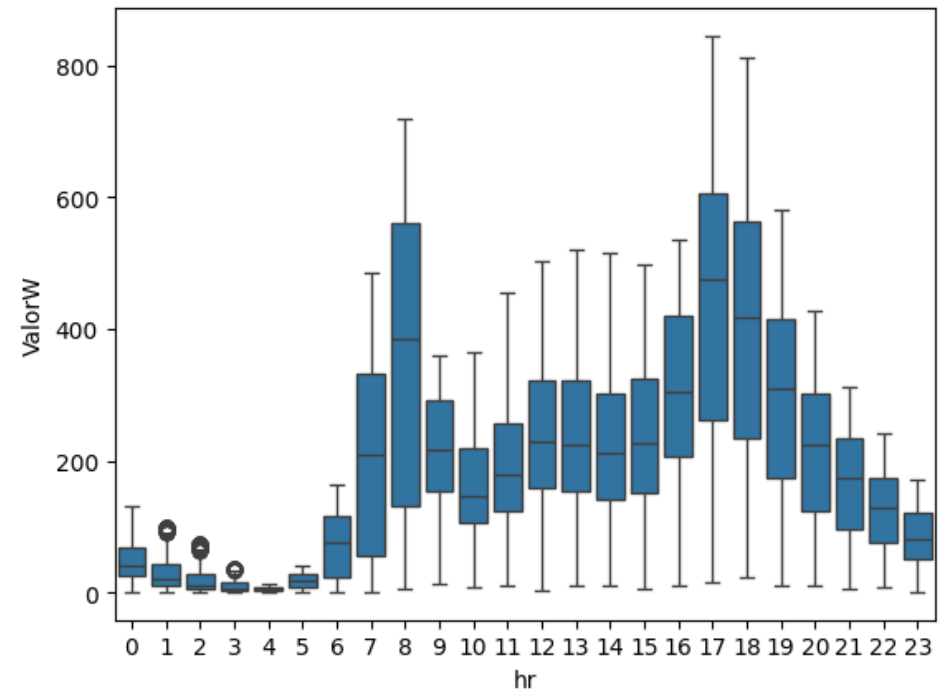
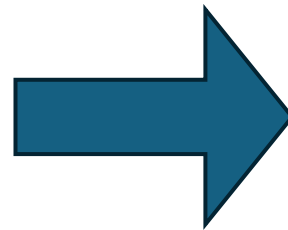
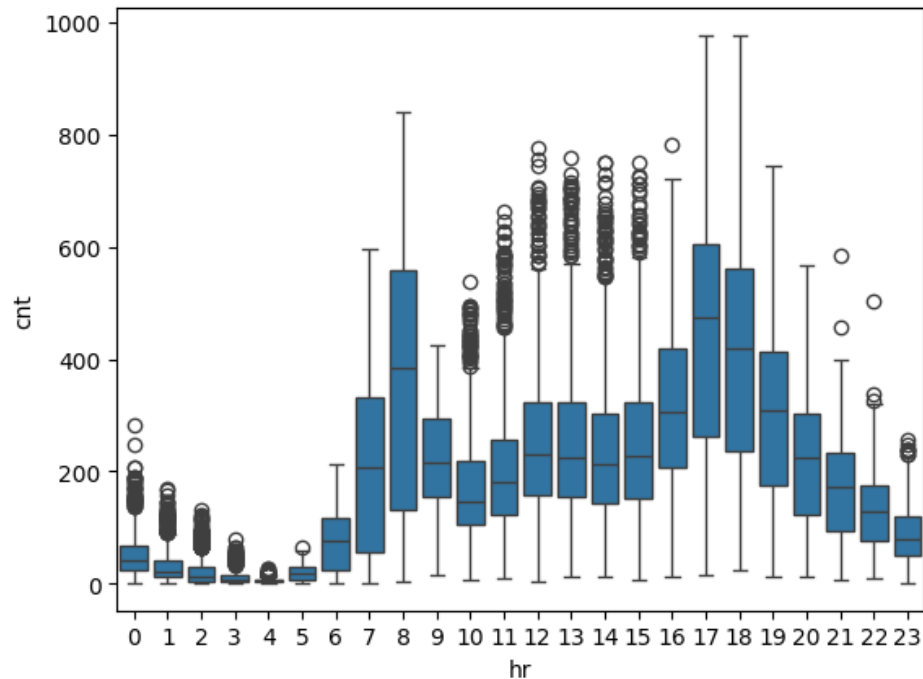
- **Depender de una línea de tiempo**

- **Ser útil**

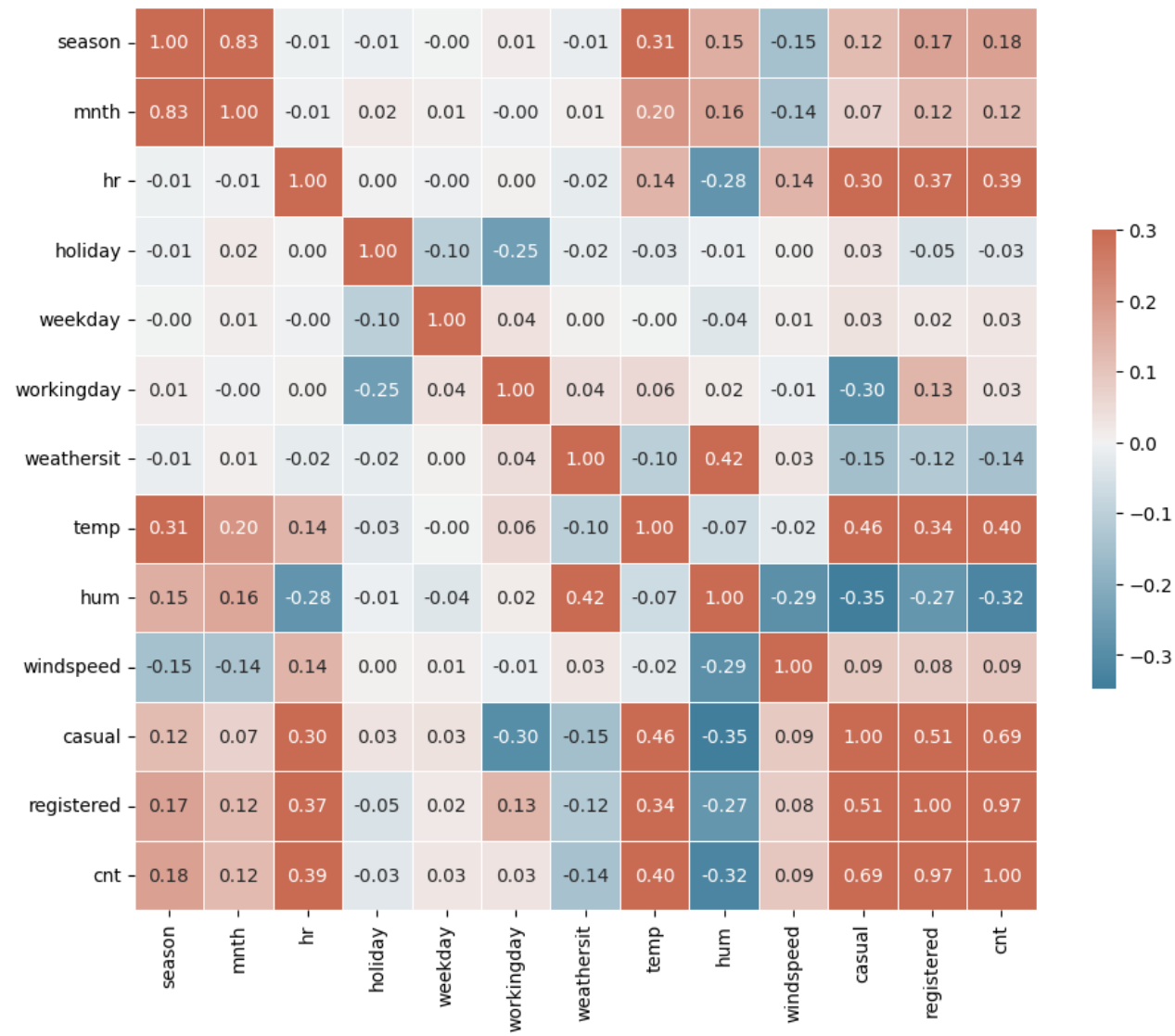
- **Instante:** índice de registro
- **Día:** fecha
- **Temporada:** estación (1:primavera, 2:verano, 3:otoño, 4:invierno)
- **Año:** año (0:2011, 1:2012)
- **Mes:** mes (1 a 12)
- **Hora:** hora (0 a 23)
- **Vacaciones:** si el día es festivo o no (extraído de [Enlace web])
- **Día laborable:** día de la semana
- **Día laborable:** si el día no es fin de semana ni festivo, es 1; de lo contrario, es 0.
- **Tiempo:** 1: Despejado, Pocas nubes, Parcialmente nublado, Parcialmente nublado 2: Niebla + Nublado 3: Nieve ligera, Lluvia ligera 4: Lluvia intensa + Paletas de hielo + Tormenta + Niebla, Nieve + Niebla
- **Temp:** Temperatura normalizada en Celsius. Los valores se calculan mediante $(t-t_{\text{mín}})/(t_{\text{máx}}-t_{\text{mín}})$, $t_{\text{mín}} = -8$, $t_{\text{máx}} = +39$.
- **atemp:** Temperatura normalizada en Celsius. Los valores se calculan mediante $(t-t_{\text{mín}})/(t_{\text{máx}}-t_{\text{mín}})$, $t_{\text{mín}} = -16$, $t_{\text{máx}} = +50$ (solo en escala horaria).
- **hum:** Humedad normalizada. Los valores se dividen entre 100 (máx.).
- **windspeed:** Velocidad del viento normalizada. Los valores se dividen entre 67 (máx.).
- **casual:** Número de usuarios ocasionales.
- **registered:** Número de usuarios registrados.
- **cnt:** Número total de bicicletas de alquiler, incluyendo tanto ocasionales como registradas

Lecciones del análisis descriptivo

- Los valores atípicos no siempre son errores de medición
- Muchas variables pueden afectar los patrones
- Matrices de correlación



Lecciones del análisis descriptivo



Análisis descriptivo



- Carga de datos
- Exploración de datos
- Inspección de valores nulos, duplicados y limpieza
- Revisión de la estadística básica
- Generación de Histogramas
- Búsqueda de valores atípicos
- Matriz de correlación
- Tratamiento de valores atípicos

matplotlib

pandas



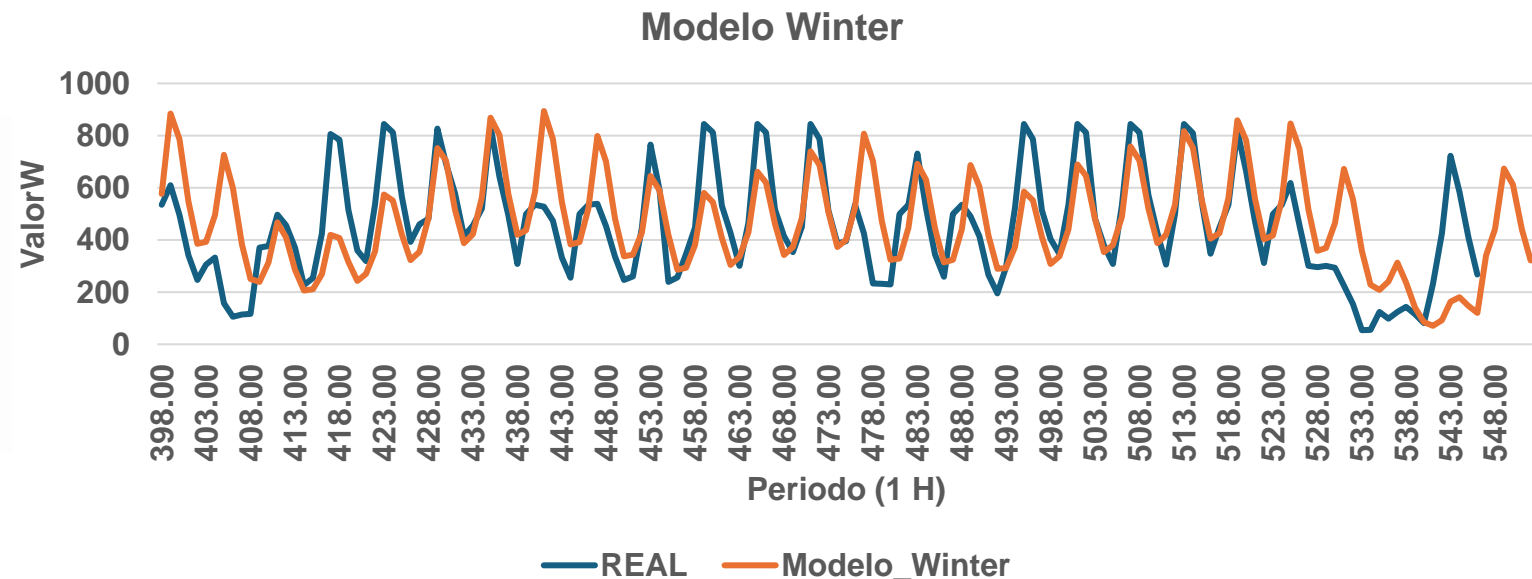
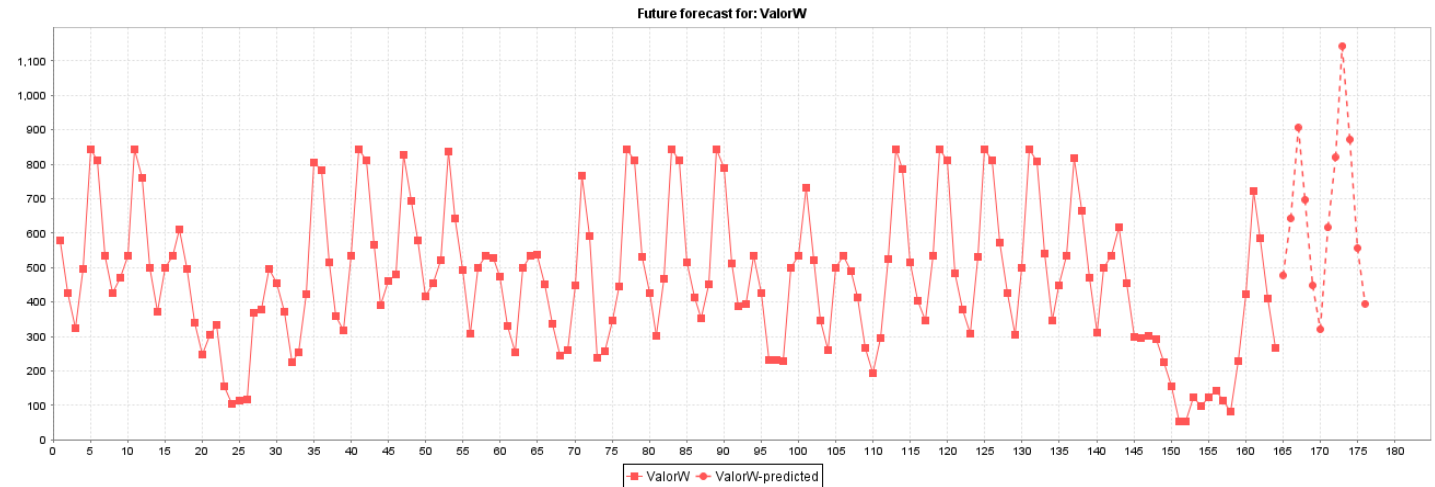
seaborn

NumPy

SciPy

Análisis predictivo

- Transformación de datos para complementar requisitos de weka.
- Definir un horizonte de búsqueda.
- Segregar los intervalos de mayor importancia para el estudio.
- Ejecución de Modelo Holt winter
- Ejecución de Modelo Holt
- Modelo Winter
- Suavización exponencial simple
- Promedio móvil



Conclusiones



- El método predictivo que presenta menor error en ejecución fue el modelo de Winter.
- Los valores atípicos tienen mucha influencia sobre el desempeño de los modelos predictivos.
- Cuando hay mucha variabilidad a lo largo de un period es recomendable estudiar el problema en bloques.

MÉTODO	MAD	MAPE (%)	RANGO TS INF	RANGO TS SUP
PROMEDIO MÓVIL (EXCEL)	151	40	-208.71	5353.03
SUAVISACIÓN EXPONENCIAL (EXCEL)	156	41	-4.63	11.42
MODELO DE HOLT (EXCEL)	161	39	-2.91	6.20
MODELO DE WINTER (EXCEL)	114	33	-12.38	22.64
MODELO HOLT-WINTER (WEKA)	135	34		