



# Anàlisis i Regressió de Dades

---

Daniel Alcover Neto - 1562889

Alexandre Galvany Pardo - 1566586

Jordi González González - 1526878



# Índex de Continguts

---

Anàlisi General de la Base de Dades

---

Selecció Atribut Principal

---

Resultats de Correlació dels Atributs

---

Estandardització de les Dades

---

Regressions Lineals Simples

---

Regressió Lineal Multivariada

---

Descens de Gradient

---

Conclusions Finals



## Anàlisi General de la Base de Dades

### *Bike Sharing*

---

Auge de Bike Sharing a diverses ciutats del món.

Interès en veure la correlació de diferents dades per l'ús de les bicicletes.

Predicció d'una alta correlació entre els dies de la setmana i el ús de les bicicletes així com amb la situació meteorològica.

Base de dades formada per:

- 16 atributs
- 732 mostres/dies registrats



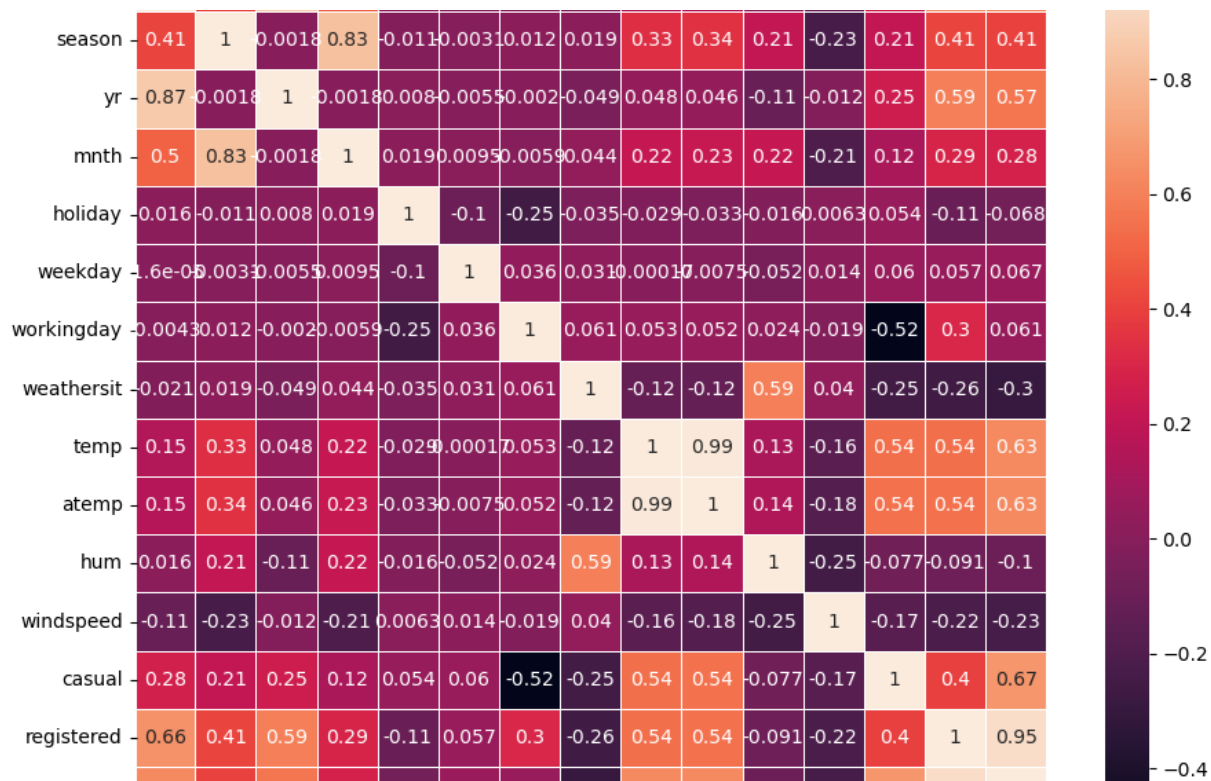
# Selecció Atribut Principal

- Atribut seleccionat: **cnt** (Numero de bicis llogades en un dia).
  - Alta correlació amb:
    - Temperatura.
    - Estació del any.
    - Usuaris registrats/ casuals.
    - Anys.
- Beneficis estudi atribut **cnt**:
  - Augmentar o disminuir la oferta segons la demanda.
  - Programar dies de manteniment de les bicicletes.
  - Fer anàlisi de l'ús en els dies segons el temps o la temperatura.
  - Comprovar quina part del any el servei és més rentable.






# Resultats correlació d'atributs




- Un cop analitzat l'histograma de correlació podem extreure la següent:
  - L'atribut amb una correlació major respecte els altres és el **cnt**, per tant hem fet una bona elecció.
- Els atributs amb més correlació del cnt són:
  - Instant (Descartat)
  - A\_temp
  - Year (Descartat)
  - Season (Descartat)
  - Casual
  - Registered

# Estandardització de les Dades

Normalització de les dades restants no normalitzades de sèrie.

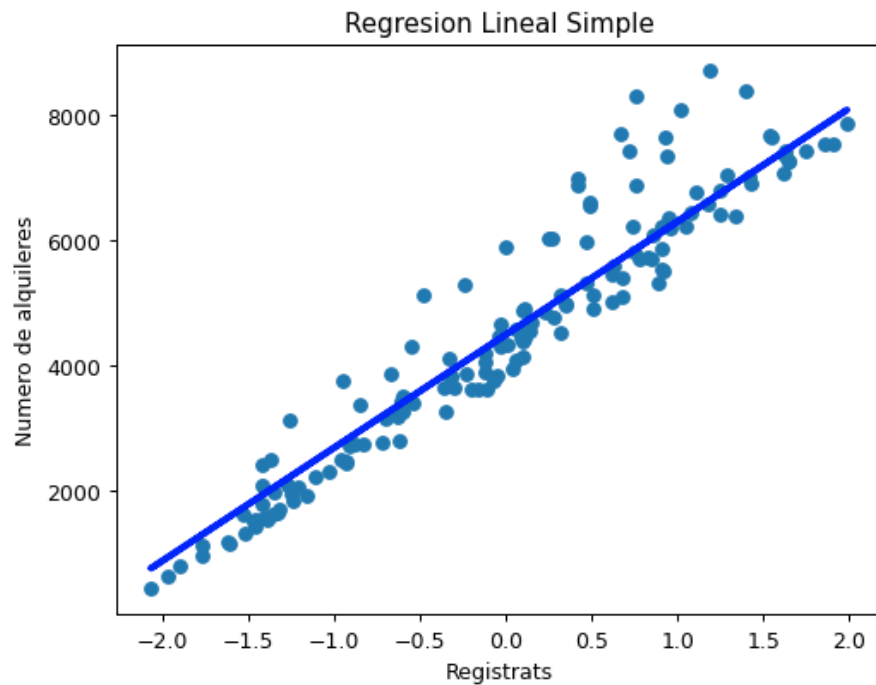


Eliminació d'un atribut que no aportava a la predicció.

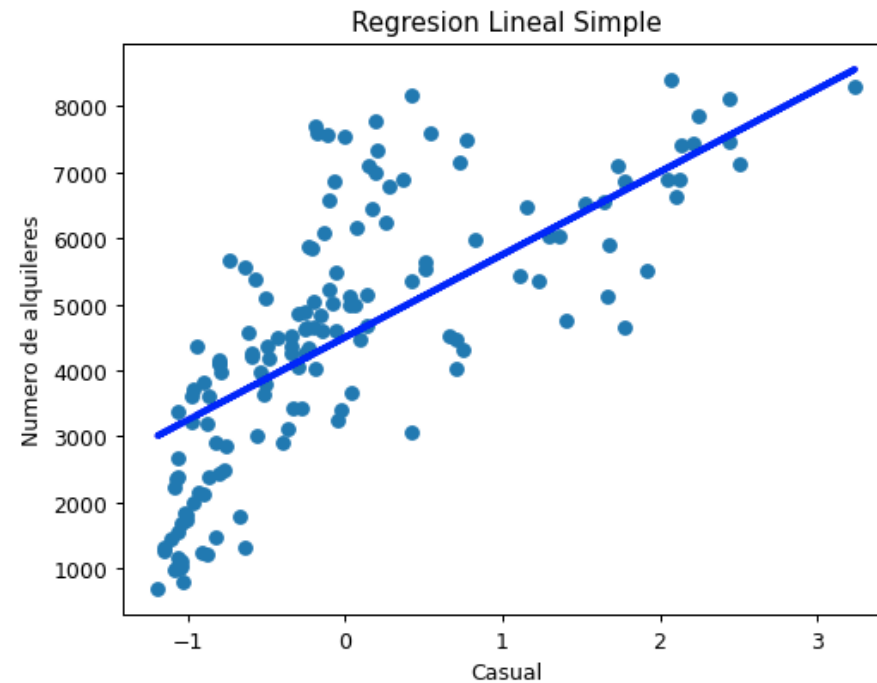


Comprovació de valors no nulls.

# Regressions Lineals Simples

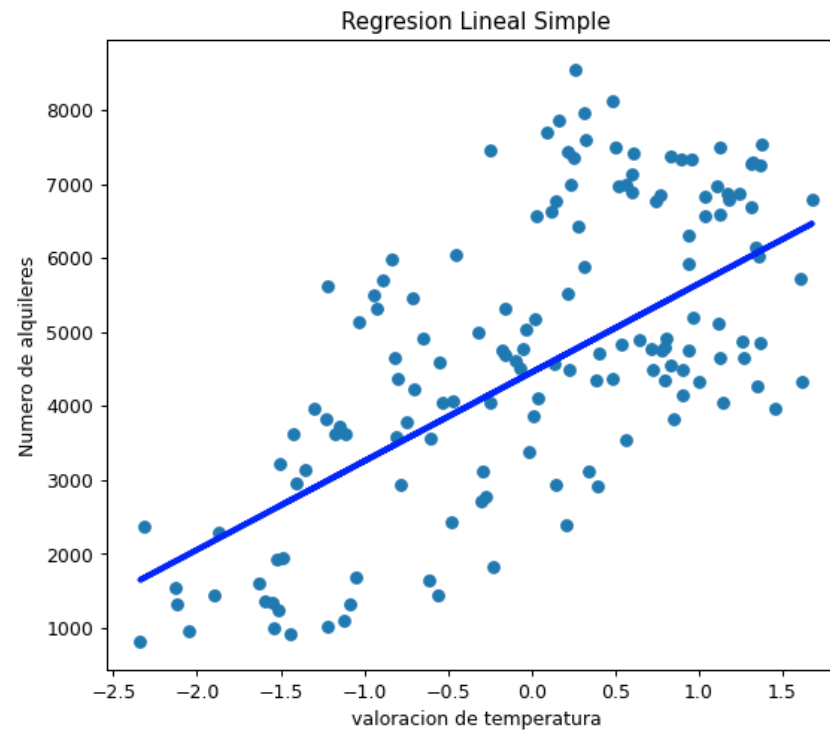


**MSE:** 568273  
**R2\_Score:** 0.910



**MSE:** 1952149  
**R2\_Score:** 0.364

# Regressions Lineals Simples



**MSE:** 2421689

**R2\_Score:** 0.378



# Regressió Lineal Multivariada

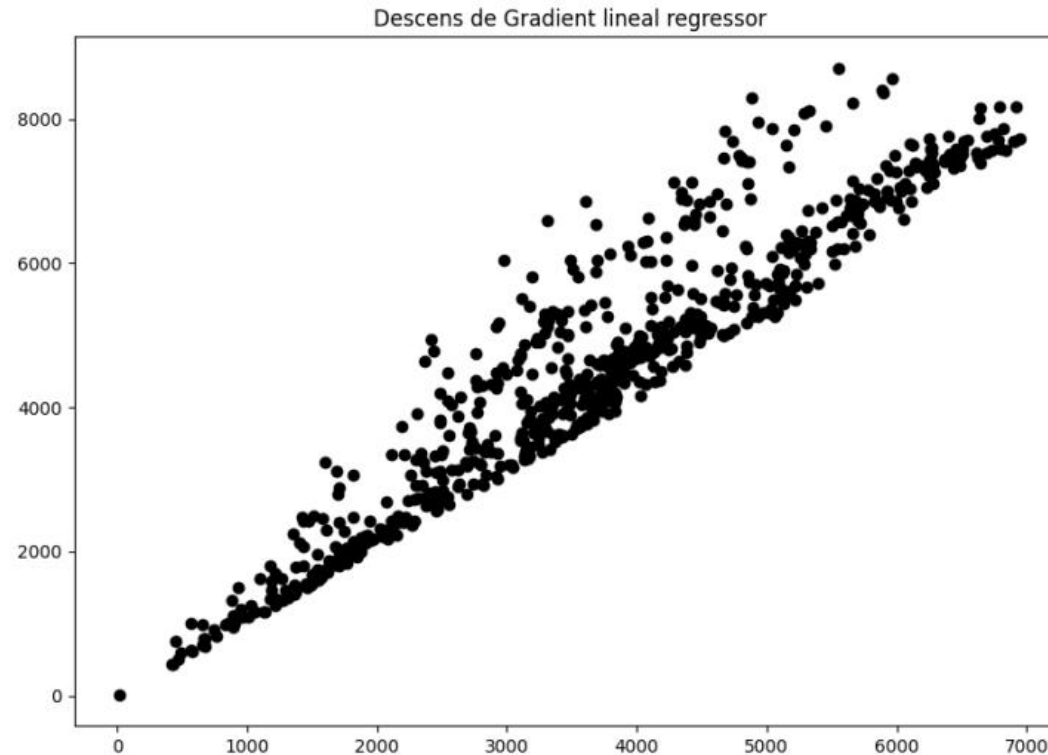
- **Definició:** Combinació de tots els atributs seleccionats en l'eix de les x respecte el nostre atribut principal que és el **cnt**.
- El resultat que hem obtingut és:

**MSE: 10406.75**

**R2\_Score: 0.9969**



# Descens de Gradient



# Conclusions Finals

---

Alta correlació atributs cnt – registred.

---

Fallida en les prediccions inicials dels atributs amb més correlació.

---

Gran importància sensació tèrmica i el ús de les bicicletes.

---

Poca correlació amb els dies de la setmana i el lloguer de les bicicletes.