Logotipo, nombre de la empresa

Descripción generada automáticamente

Universidad de Montevideo – Introducción a la Ciencia de Datos

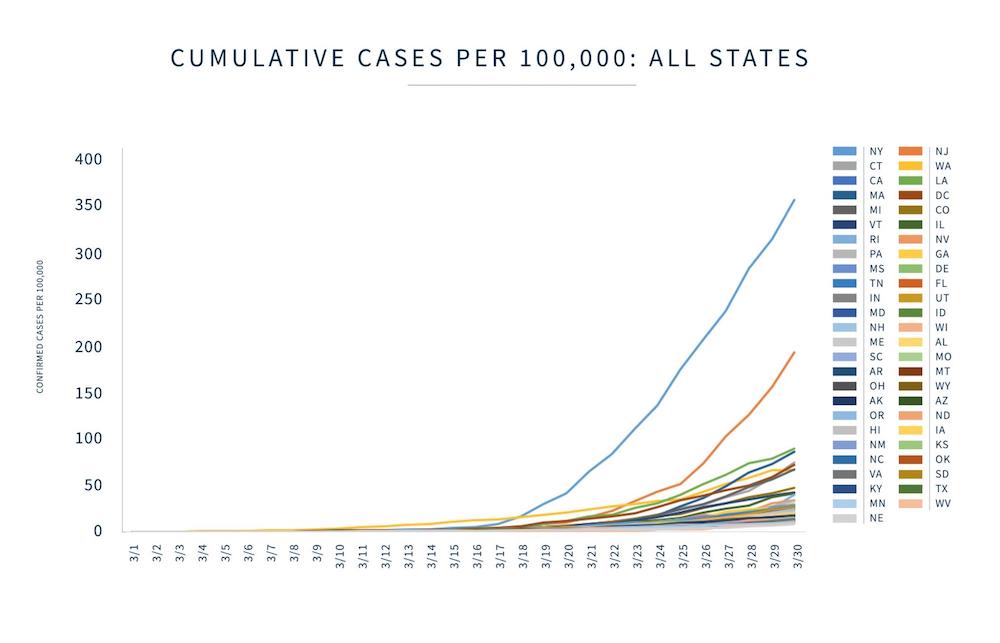
Examen 2021 - S1

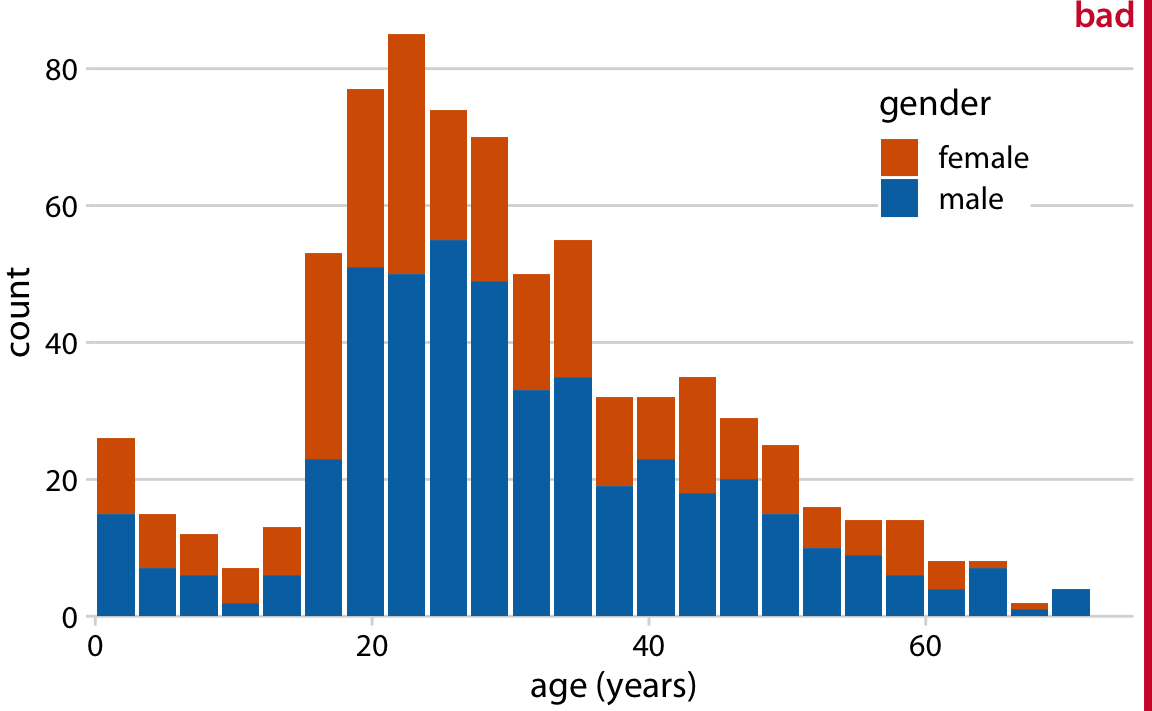
5 de Julio, 2021

1. Explique el método de validación k-fold cross-validation haciendo un diagrama e indicando el propósito de cada dataset.
2. Usted entrena un modelo predictivo con un conjunto de training y evalúa la performance utilizando una métrica. Luego, evalúa la performance en un conjunto de testing y nota una gran diferencia de performance (mucho menor performance en testing en relación a training). Qué problema es este? Sugiera al menos una manera de resolverlo.
3. Explique al menos 4 pre-procesamientos utilizados en NLP.
4. Explique al menos 3 métodos de feature engineering para datos tabulares.
5. Corredores de distintos países corren 10 carreras y obtienen los tiempos que se muestran en la tabla del anexo. La desviación estándar para los corredores del país 1 es de 103 segundos mientras que para los corredores del país 2 es de 101 segundos. Estableciendo un nivel de significancia alpha = 0.05, existe una diferencia significativa en la media de tiempos de ambos corredores?

La tabla del anexo posee valores negativos del z-score porque es de la cola izquierda, pero es igual para los valores positivos del z-score si fuera la cola derecha.

1. Comente los errores que existen en las gráficas presentes en el anexo.
2. En clase vimos grandes categorías de tareas relacionadas a Data Science y Machine Learning. Estas eran: supervised learning, unsupervised learning y reinforcement learning. Explique con sus palabras estas categorías de tareas.
3. Explique 3 de los siguientes conceptos relacionados a una serie de tiempo: tendencia, estacionalidad, cambio cíclico y cambio irregular. Explique además como se reflejan dichos conceptos en la gráfica del anexo.
4. Explique un problema que podría ser resuelto utilizando unsupervised learning, citando los datos que utilizaría y qué algoritmo.
5. Usted quiere usar una regresión lineal para generar predicciones de ventas. Qué hipótesis asume el modelo lineal? Cómo se interpreta el modelo lineal? Explique los conceptos de pendiente, intercepción y R2.

Problema 6:  
  




A screenshot of a computer

Description automatically generated with medium confidence

Problema 5:

Table

Description automatically generated

Table

Description automatically generated

Problema 8:

Chart, line chart

Description automatically generated