

## Examen Transversal

Sigla	Nombre Asignatura	horas semana
MDY7101	Minería de datos	5 h/semana 18

Ítem	Puntaje	% Ponderación
Competencia Especialidad	42 puntos	100%

### INSTRUCCIONES GENERALES:

#### 1. DESARROLLO DEL PRODUCTO

- 1.1. El Examen Transversal de **Minería de Datos**, comprende todo el ciclo desde el análisis de la información hasta definir un modelo adecuado para aplicar a los datos solicitados.
- 1.2. El Examen es grupal de hasta 3 alumnos.
- 1.3. El examen se entregará en la semana 4 a los estudiantes.
- 1.4. En la semana 17-18 los alumnos deberán presentar el trabajo final con su documentación del semestre y además una presentación explicando los resultados obtenidos.
- 1.5. El tiempo de la presentación es de 15 min.
- 1.6. El estudiante deberá realizar el análisis del caso asignado, el cual corresponde a un trabajo semestral en donde a medida que va cumpliendo las unidades de estudio deberá ir avanzando en el.
- 1.7. El examen deberá realizarse considerando los siguientes aspectos:
  - 1.7.1. Exploración, limpieza y transformación de datos
  - 1.7.2. Desarrolla un proyecto de minería de datos utilizando la herramienta Jupyter Notebook.
  - 1.7.3. Realice un análisis de la organización y de los datos que disponen antes de comenzar a trabajar los registros.
  - 1.7.4. Realizar el análisis exploratorio de los datos (EDA), obtener representación gráfica de estos para tomar decisiones acertadas sobre la utilidad de los datos conforme al comportamiento observado.
  - 1.7.5. Generar información y conocimiento útil desde las fuentes de datos y la experiencia, para apoyar la toma de decisiones de acuerdo con las necesidades de la organización.
  - 1.7.6. Evalúe usando algoritmos supervisados y no supervisados.
  - 1.7.7. Evalúa el algoritmo en base a las métricas que corresponda para determinar cuál es el más adecuado al contexto.

# Examen Transversal

---

## 2. ENTREGA DEL PRODUCTO

La entrega del producto final será en la semana 17-18.

El informe y presentación deben contener los siguientes puntos:

- Descripción del contexto
- Propósito y justificación del proyecto
- Realizar análisis de los datos disponibles y encontrar patrones de comportamiento.
- Explicar paso a paso lo realizado y adjuntar el notebook generado por la herramienta Jupyter Notebook con el código Python usado para el manejo de datos.
- Realizar razonamiento y justificación respecto del análisis de los datos
- Definición de fases, tareas y entregables a partir del dataset y el proyecto en general.
- Sugerir decisiones estratégicas para apoyar la toma de decisiones que ayudarán a la organización.

Formatos de entrega:

- Hoja tamaño carta o A4
- Tipo de letra: Títulos Arial 14 Negrita, Contenido Arial 12
- Interlineado: 1,5.
- Párrafo: Justificado.

NOTA: Dentro de la evaluación del examen serán considerados los siguientes puntos:

- Formato Presentación Informe.
- Ortografía.
- Redacción.

# Examen Transversal

## Caso de Estudio

Australia es uno de los países con mayor extensión del mundo y el continente más seco y llano del planeta. Aproximadamente la mitad occidental de Australia es una extensa meseta árida. El clima predominante es el desértico y semiárido (se estima que un 40% del territorio lo forman dunas de arena), si bien el norte del país cuenta con un clima tropical. Las tierras más fértiles y el clima templado con cuatro estaciones se concentran en el sureste y suroeste. La principal cordillera del país, la Gran Cordillera Divisoria (The Great Dividing Range), se extiende a lo largo de más de 3.500 km entre los estados de Queensland y Victoria. El pico más alto es el Monte Kosciuszko con 2.228 metros. Los ríos más caudalosos nacen en la Gran Cordillera Divisoria y, en su gran mayoría, fluyen hacia el este y desembocan en el océano Pacífico. Por su longitud y caudal destacan dos ríos: el Murray y el Darling. Australia se divide en seis estados (New South Wales, Victoria, Queensland, South Australia, Western Australia y Tasmania) y dos territorios continentales (Northern Territory y Australian Capital Territory (ACT)). La ciudad de Sídney, con sus 5,73 millones de habitantes, es la más poblada de Australia y es la capital del estado de Nueva Gales del Sur. Es a su vez un importante centro industrial y puerto comercial de primera magnitud. La ciudad es muy extensa, con un radio aproximado de 40 km, incluyendo los suburbios. Otras ciudades importantes son Melbourne (5,19 millones), Brisbane (2,56 millones), Perth (2,38 millones) y Adelaide (1,41 millones). Casi el 90% de la población se concentra en las zonas urbanas. De sus más de 25 millones de habitantes, un 65,5% se encuentra en edad de trabajar (de 15 a 64 años), mientras que aquellos con más de 65 años y con menos de 15 años suponen el 15,7% y el 18,8% de la población respectivamente. La superficie del país es de aproximadamente 7.741.220 km<sup>2</sup>, quince veces superior a España, representando el 5% del total de la superficie terrestre emergida. Es el sexto país del mundo en extensión. La distancia máxima del continente australiano de este a oeste es de 4.100 km y de norte a sur, de 3.200 km. La longitud costera es de 36.735 km. Su vasta extensión inevitablemente condiciona el tráfico de personas, mercancías y servicios. Australia es, además, el más llano de los continentes, con una altitud media de menos de 300 metros, siendo The Great Dividing Range la única gran cordillera de importancia. Casi dos tercios del territorio carecen de corrientes de agua hacia el mar. La principal cuenca fluvial es la del Murray-Darling con 1.061.469 km<sup>2</sup>, que drena las vertientes sur occidentales de la Gran Cordillera. El país cuenta con una gran diversidad climática que abarca desde el clima tropical en el norte, que representa el 39% del territorio, al clima templado-continental en el sudeste y Tasmania. En las zonas del centro predomina el clima desértico. La zona más fértil es la franja costera entre Sídney y Adelaide, con lluvias moderadas todo el año. Australia está menos sujeta a climas extremos que otros países en su misma latitud, debido a los efectos moderadores de los mares y océanos circundantes y a la ausencia de grandes montañas. Las estaciones varían con la latitud pero aproximadamente son: primavera (septiembre y octubre), verano (noviembre-marzo), otoño (abril y mayo) e invierno (junio-agosto). Las temperaturas medias oscilan entre los 27°C en la zona norte y los 13°C en las zonas más al sur, alcanzando las máximas en la zona centro (38°C). Por tanto, las principales características climáticas australianas son inviernos suaves y veranos cálidos, así como abundante sol y poca

## Examen Transversal

humedad. Las precipitaciones son muy escasas en el interior y aumentan en las zonas costeras, de modo que las zonas mejor regadas son los litorales norte, este, sudeste y sudoeste. En el norte del país hay dos estaciones, seca en invierno y húmeda en verano con la irrupción de lluvias monzónicas. Australia es, después de la Antártida, el continente más seco. Aun así, y debido a su diversidad climática, se dan todo tipo de fenómenos naturales extremos como sequías, inundaciones, ciclones tropicales, vendavales, incendios forestales (en lo que se conoce como bush o monte en Australia) y, ocasionalmente, tornados. En cuanto a recursos naturales y minería, Australia es uno de los principales productores y 3 exportadores de minerales y productos energéticos a nivel mundial. El sector minero representa alrededor del 10% del PIB con un valor de 148.000 millones de AUD. En 2018-19 presentó una tasa de crecimiento del 6%. El valor de las exportaciones totales de productos mineros y energéticos se espera que alcance los 299.000 millones AUD en el periodo 2019-20. Las exportaciones de mineral de hierro alcanzaron durante 2018-19 los 100.000 millones de AUD. Australia está entre los cinco principales países exportadores del mundo de bauxita, alúmina, mineral del hierro, zinc, carbón y de gas natural licuado (GNL).

Texto extraído del sitio <https://www.icex.es/> ministerio de industria comercio y turismo de España.

Se dispone de un set de datos de observaciones meteorológicas diarias de múltiples ubicaciones en Australia, obtenidas de la Oficina de Meteorología de la Commonwealth de Australia y procesadas para crear este conjunto de datos de muestra, los datos se han procesado para proporcionar una variable objetivo RainTomorrow (si hay lluvia al día siguiente - No / Sí) y una variable de riesgo RISK\_MM (cuánta lluvia registrada en milímetros), esta información se ha dejado disponible para que usted la explore y busque información que sea relevante para demostrar su aprendizaje de Minería de datos.

### El dataset posee los siguientes campos:

Columna	Descripción
Fecha	fecha de la observación
Ubicación	ubicación de la estación meteorológica
MinTemp	temperatura mínima en grados Celsius
MaxTemp	temperatura máxima en grados Celsius
Lluvia	cantidad de lluvia registrada ese día en mm.
Evaporacion	evaporación (mm) en 24 horas
Sol	número de horas de sol brillante en el día
DirRafaga	dirección de la ráfaga de viento más fuerte en 24 horas.
VelRafaga	velocidad (km/hr) de la ráfaga de viento más fuerte en 24 horas.
Dir9am	dirección del viento a las 9am
Dir3pm	dirección del viento a las 3pm
Vel9am	velocidad (km/hr) del viento a las 9am

## Examen Transversal

Vel3pm	velocidad (km/hr) del viento a las 3pm
Hum9am	porcentaje de humedad a las 9am
Hum3pm	porcentaje de humedad a las 3pm
Pres9am	presión atmosférica (hpa) a nivel del mar a las 9am
Pre3pm	presión atmosférica (hpa) a nivel del mar a las 3pm
Nub9am	fracción del cielo cubierto por nubes a las 9am. Se mide en "octavos", de manera que un valor 0 indica cielo totalmente despejado y 8, cielo totalmente cubierto.
Nub3pm	fracción del cielo cubierto por nubes a las 3pm. Se mide en "octavos", de manera que un valor 0 indica cielo totalmente despejado y 8, cielo totalmente cubierto.
Temp9am	temperatura en grados celsius a las 9am
Temp3pm	temperatura en grados celsius a las 3pm
LluviaHoy	variable indicadora que toma el valor 1 si la precipitación en mm. en las últimas 24 hrs. excede 1 mm. y 0 si no.
RISK_MM	La cantidad de lluvia. Una especie de medida del "riesgo".
LluviaMan	variable indicadora que toma el valor 1 si al día siguiente llovió y 0 si no.