Prueba Técnica para Científico de Datos

naowee

Metodología

Lea con detenimiento los requerimientos y condiciones para la ejecución de la prueba técnica.

naowee

Requerimientos técnicos

Se evaluará las habilidades y conocimientos para el análisis de datos desde la manipulación, transformación y análisis de datos a través de dos casos prácticos.

Así mismo, se evaluará la capacidad de análisis de negocio desde el enfoque que se tome para resolver cada caso.

Realice la prueba utilizando Python.

Debe entregar la prueba así:

1) Si usa Google Colab: link de Google Colab + Link GitGist.
2) Si usa Script Python: Repositorio GitHub con los archivos .py utilizados.
3) Debe entregar los archivos que utilice de apoyo.

Fecha límite de entrega: Tendrá disponibilidad para entregar la prueba hasta el martes 9 de diciembre a las 8:00 a.m.

Todo lo desarrollado durante la prueba debe ser sustentado en la entrevista técnica.

naowee

Caso 1

Resultados Mundiales de Fútbol Femenino

naowee

Caso: Análisis de la Copa Mundial Femenina

Contexto y Desafío:
La Copa Mundial Femenina de la FIFA ha experimentado un crecimiento exponencial desde su primera edición en 1991, atrayendo una audiencia global y consolidando el fútbol femenino como un pilar del deporte internacional. Con la evolución del torneo, ha surgido la necesidad de analizar tendencias en el desempeño de las selecciones, patrones de juego y factores que contribuyen al éxito en la competición.

El Reto Analítico:
La FIFA está interesada en un análisis detallado del rendimiento de los equipos en las distintas ediciones del torneo, identificando patrones en la cantidad de goles, frecuencia de victorias y desempeño a lo largo de los años. Se busca descubrir qué selecciones han dominado históricamente y cómo ha evolucionado el nivel de competitividad.

Metodología y Enfoque:
A través del análisis de datos de partidos de la Copa Mundial Femenina desde 1991 hasta 2023, se espera:
•Identificar tendencias de goles y resultados a lo largo de los torneos.
•Determinar los equipos con mayor rendimiento en cada edición.
•Explorar patrones de juego y evolución del fútbol femenino a nivel internacional.

Preguntas Clave:
•¿Cómo ha cambiado el promedio de goles por partido a lo largo de los torneos?
•¿Cuáles son las selecciones con mejor desempeño en términos de victorias?
•¿Existen tendencias en los equipos dominantes y con peor desempeño?

naowee

Caso: Análisis de la Copa Mundial Femenina

Con base en los dataset world_cup_women:
https://raw.githubusercontent.com/daramireh/simonBolivarCienciaDatos/refs/heads/main/world_cup_women.csv

y world_cup_matches:
https://raw.githubusercontent.com/daramireh/simonBolivarCienciaDatos/refs/heads/main/matches_1991_2023.csv realice:

1. Análisis donde identifique las variables del conjunto de datos mostrando los valores nulos, duplicados y el tipo de variable.

2. Validación cruzada entre las tablas, identifique los campos que relacionan las tablas y si existen datos faltantes en dichos campos.

3. Elabore la tabla de posiciones del mundial realizado en 1991. Tenga en cuenta que cada partido ganado da 3 puntos, cada partido empatado da 1 punto. Así mismo, las tarjetas amarillas suman -1 punto para juego limpio y las tarjetas rojas – 2 puntos a juego limpio.
   La tabla debe tener la siguiente estructura:
   Equipo | Partidos Jugados (PJ) | Partidos Ganados (PG) | Partidos Empatados (PE) | Partidos Perdidos (PP)| Goles a Favor (GF) | Goles en Contra (GC) | Diferencia de Goles (GF – GC) | Juego Limpio (JL) | Puntos

4. Elabore la tabla de goleadoras del mundial de 2023.

naowee

Caso: Análisis de la Copa Mundial Femenina

5. Construya una rutina cuya salida sea UNA ÚNICA TABLA que muestre:

Año | Host | Equipo | Partidos Jugados | Goles Totales marcados | Promedio de Goles marcados | Goles Totales recibidos | promedio de goles recibidos | partidos totales ganados | partidos totales perdidos | partidos totales empatados | promedio de asistencia por equipo.

naowee

Caso 2

Desempeño estudiantes de matemáticas

naowee

Caso: Factores que Impactan el Desempeño en Matemáticas

El Reto Analítico:
Una universidad ha recopilado información sobre el desempeño de sus estudiantes en matemáticas con el objetivo de identificar los factores que más influyen en su rendimiento. Se busca entender cómo aspectos como el tiempo de estudio, la asistencia a actividades extracurriculares y las horas de sueño inciden en el puntaje final.

Metodología y Enfoque:
A través del análisis de datos de estudiantes universitarios, se espera:
•Evaluar la relación entre el tiempo de estudio y las calificaciones finales.
•Identificar diferencias de desempeño según la asistencia a actividades extracurriculares.
•Aplicar modelos de regresión para predecir el rendimiento académico y entender sus principales determinantes.

Preguntas Clave:
•¿Cuáles son los principales factores que afectan el rendimiento en matemáticas?
•¿Existe una correlación entre el tiempo de estudio y la calificación final?
•¿Cómo varía el desempeño según la asistencia a actividades extracurriculares y la resolución de los ejercicios prácticos propuestos en el material de estudio?

naowee

Caso: Factores que Impactan el Desempeño en Matemáticas

Con base en el dataset https://raw.githubusercontent.com/daramireh/simonBolivarCienciaDatos/refs/heads/main/Student_Performance.csv realice:

1. Análisis de Estructuras de Datos a través de:
    a. Visualizaciones de datos: es importante que usted haga una descripción de los hallazgos.
    b. Identifique las variables, tipo de dato y elabore un diccionario de datos con estas variables (puede hacerlo en Excel conectándose directamente al Raw).
    c. ¿Es necesario transformar los datos para mejor entendimiento de los mismos? Si su respuesta es positiva, ¿qué tipo de transformaciones haría? Si considera necesario, realice las transformaciones para desarrollar los puntos posteriores.

2. Como resultado del análisis de las estructuras de datos del punto anterior, elabore un análisis exploratorio donde identifique:

    a. Estadística descriptiva de las variables. Haga una descripción de los hallazgos.
    b. ¿Existen diferencias estadísticas entre el puntaje final (Performance Index) y la asistencia a actividades extracurriculares? En la resolución se debe poder observar todo el proceso realizado para la respuesta.
    c. Identifique los grupos o conglomerados de estudiantes según sus características.

naowee

Caso: Factores que Impactan el Desempeño en Matemáticas

Es necesario identificar potenciales estudiantes con bajo desempeño por lo que es importante desarrollar un producto de datos que le facilite a la institución la identificación temprana de estos casos. Para ello, le han solicitado a usted desarrollador una API con las siguientes condiciones:

1. Debe desarrollar un modelo predictivo utilizando Machine Learning de los estudiantes con bajo rendimiento académico (defina usted como los identificaría a partir del análisis exploratorio hecho anteriormente).

Nota: Para el entregable de la prueba debe entrenar un modelo de regresión y un modelo de clasificación.

2. Con los modelos entrenados, elija uno el cuál debe ser ser comparado con al menos 2 modelos más para comprobar que es el más eficiente para identificar el bajo desempeño académico.

3. Debe definir la métrica de desempeño utilizada para seleccionar el mejor modelo.

4. Debe desarrollar una API para poder consumir el modelo entrenado. La API debe ser desarrollada en Python preferiblemente con FastAPI y dockerizada. Debe entregar el link del repositorio con la API.

    1. Si conoce los principios SOLID aplíquelos.
    2. Simule el CRUD de la API con diccionarios.
    3. Entregue las muestras utilizadas comprobar el funcionamiento de la API ya sea en Postman o en Swagger.
    4. Si no sabe usar Docker, elabore la API con un entorno virtual.

Tel: (+57) 300 2833285
Dirección: Cra. 55 #100 - 51
Centro empresarial Blue Gardens - Oficina 608
Contáctanos: www.naowee.com