Indice

1. *Descripción del dataset (0,5 puntos) 2*
2. *Integración y selección de los datos a analizar (0,5 ptos) 4*
3. *Limpieza de datos (2 ptos..... 4*
4. *Análisis de los datos (2,5 ptos)..... 4*
5. *Representación de los resultados (2 ptos) 4*
6. *Resolución del problema (0,5 ptos)..... 4*
7. *Resolución del problema (2 ptos)..... 4*

Referencias Bibliográficas:..... 5

1. Descripción del dataset (0,5 puntos)

Para la realización de esta práctica se utilizará el dataset generado en la PRA1. Se trata de un dataset que recopila los datos de los sorteos de la lotería primitiva que se han celebrado desde el jueves 21 de junio de 1991 hasta el 31 de octubre de 2020. Es un dataset en formato .csv (Historico_Primitiva.csv)

Los campos del fichero Historico_Primitiva.csv son:

Nombre Campo	Tipo	Descripción	Ejemplo
INDICE	Numérico	Número de línea	1
FEC_LARGA	Texto	Fecha descriptiva	Resultado del sábado 31 de octubre de 2020
FEC_SORTEO	Fecha	Fecha del sorteo en formato AAAMMDD	20201031
B1	Numérico	Valor Bola 1	2
B2	Numérico	Valor Bola 2	13
B3	Numérico	Valor Bola 3	21
B4	Numérico	Valor Bola 4	32
B5	Numérico	Valor Bola 5	41
B6	Numérico	Valor Bola 6	43
C	Numérico	Complementario	40
R	Numérico	Reintegro	9
RECAUDACION	Numérico	Importe recaudado	11.420.804,00 €
BOTE	Numérico	Importe del bote	7.500.000,00 €
ACE	Numérico	Acertantes de la categoría especial (6 aciertos + R)	0
PCE	Numérico	Premio de la categoría especial	0,00 €
AC1	Numérico	Acertantes de la categoría 1ª (6 aciertos)	1
PC1	Numérico	Premio categoría 1ª	1.450.282,32 €
AC2	Numérico	Acertantes de la categoría 2ª (5 aciertos + C)	9

PC2	Numérico	Premio categoría 2ª	24.171,37 €
AC3	Numérico	Acertantes de la categoría 3ª (5 aciertos)	206
PC3	Numérico	Premio categoría 3ª	2.288,07 €
AC4	Numérico	Acertantes de la categoría 4ª (4 aciertos)	10.098
PC4	Numérico	Premio categoría 4ª	75,40 €
AC5	Numérico	Acertantes de la categoría 5ª (3 aciertos)	189.207
PC5	Numérico	Premio categoría 5ª	8,00 €
ACR	Numérico	Acertantes del Reintegro	1.100.069
PCR	Numérico	Premio Reintegro	1,00 €

Sobre este dataset haremos diversos estudios estadísticos descriptivos:

- Histograma de la frecuencia de los números aparecidos
- Importe en premios por año
- Media de la recaudación
- …

e intentaremos responder a preguntas como:

Teniendo en cuenta que en ese periodo de tiempo hemos sufrido dos grandes crisis (la del 2008 y la actual producida por la pandemia del COVID-19)

¿Existe alguna relación entre la serie temporal de la recaudación de los años 2000 al 2020 y los índices económicos en ese periodo de tiempo? Es decir ¿La recaudación aumenta en tiempos de crisis o disminuye?

Para ello además contaremos con el fichero "variacion_del_pib.csv" en el que se presenta la variación anual del PIB de España hasta el tercer trimestre de 2020 extraído de:

https://www.epdata.es/datos/pib-espana-ine-contabilidad-nacional-trimestra/36/espana/106

Nota: Está pendiente pensar qué otras preguntas podemos responder.

2. Integración y selección de los datos a analizar (0,5 ptos)

Seleccionaremos de cada dataset los datos correspondientes a los años 2000 a 2020. Puesto que los datos del PIB son trimestrales y los de la recaudación tienen información dos veces a la semana, habrá que establecer un criterio para poder homegeneizar los intervalos.

3. Limpieza de datos (2 ptos)

Aquí vemos si existen datos en blanco o nulos y valores extremos y definimos cómo trataremos estos registros.

4. Análisis de los datos (2,5 ptos)

En este apartado realizaremos los análisis estadísticos necesarios.

5. Representación de los resultados (2 ptos)

En este apartado se muestran los histogramas descriptivos y las series temporales donde se verá visualmente la relación entre la recaudación y el PIB.

6. Resolución del problema (0,5 ptos)

Después del trabajo realizado en los puntos anteriores, podremos dar respuesta a las cuestiones planteadas al principio.

7. Código (2 ptos)

A continuación se incluye el código en Python del proceso de limpieza y análisis. (PENDIENTE)

Firma de participantes:

Contribuciones	Firma
Investigación previa	MAA
Redacción de las respuestas	MAA
Desarrollo código	MAA

Referencias Bibliográficas: