

Tarea 1 Módulo 1 - Análisis de Datos 2022-1

D. Sierra-Porta

March 25, 2022

1 Una imagen es un arreglo de datos.

Busque en internet en google, cualquier par de imágenes que usted desee, las que quiera y tengan sentido para ustedes. Una imagen es en realidad un arreglo de datos, en este caso tridimensional. Tendrá siempre un eje X y un eje Y que designan el largo y ancho de la imagen (en píxeles), y además tendrá una tercera dimensión Z la cual tiene una dimensionalidad de $d(Z) = 3$ que representa la escala de color en tonos azules, rojos y verdes (BRG, Blue-Red-Green).

Lo que queremos es hacer estadística con las componentes de los tonos de la imagen y hacer análisis descriptivo de dichas componentes. Supongamos que usted depara todos los colores azules, los verdes y los rojos. Entonces:

1. Construya un arreglo de estas componentes en un sólo (otro) arreglo de datos.
2. Calcule para cada arreglo las medidas de tendencia central, las medidas de dispersión y las medidas de simetría correspondientes, también las medida de variabilidad o coeficiente de variabilidad o coeficiente de Pearson. Diga si las medidas son robustas o no.
3. Haga un histograma de cada componente.
4. Ahora suponga que la escala de cada color, la cual es un número que va en el intervalo $[0, 255]$ usted la divide en ["Claro", "Medio", "Oscuro"], bajo un criterio que usted elabore. Construya un nuevo arreglo con estos datos.
5. Ahora haga un gráfico de torta y un gráfico de barras para visualizar el último arreglo que hizo.
6. Haga un análisis general de la situación sin dejar de mencionar cada cálculo que ha hecho que no sobrepase la dimensión de 15 líneas ni sea menor a 10 líneas.

2 Asignación - Pandas, Pandas, Pandas. ¡La inmersión en los datos continúa! Héroe de Pymoli

Ahora, es el momento de tomar lo que has aprendido sobre Python Pandas y aplicarlo a nuevas situaciones. Para esta tarea, tendrás que completar los Desafíos de Datos. Sólo asegúrate de darlo todo - ya que las habilidades que perfeccionas se convertirán en poderosas herramientas en tu cinturón de herramientas de análisis de datos.

¡Enhorabuena! Después de un duro trabajo en las minas de datos, has conseguido un trabajo como analista principal en una empresa de juegos independiente. Te han asignado la tarea de analizar los datos de su juego de fantasía más reciente, Heroes of Pymoli.

Como muchos otros de su género, el juego es free-to-play, pero se anima a los jugadores a comprar objetos opcionales que

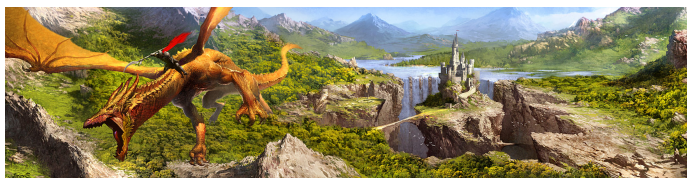


Figure 1: Heroes of Pymoli - Video juego.

mejoren su experiencia de juego. Como primera tarea, la empresa quiere que usted genere un informe que desglose los datos de compra del juego para obtener información significativa.

Tu informe final debe incluir cada uno de los siguientes datos

1. Recuento de jugadores
 - Número total de jugadores
2. Análisis de compras (total)
 - Número de artículos únicos
 - Precio medio de compra
 - Número total de compras
 - Ingresos totales
3. Datos demográficos por género
 - Porcentaje y número de jugadores masculinos
 - Porcentaje y número de jugadoras
 - Porcentaje y recuento de otros / no revelados
4. Análisis de compras (género). A continuación, cada uno de ellos desglosado por género
 - Recuento de compras
 - Precio medio de compra
 - Valor total de las compras
 - Total medio de compras por persona, por género
5. Datos demográficos por edad. A continuación, cada uno de ellos desglosado en franjas de 4 años (es decir, <10 , 10-14, 15-19, etc.)
 - Recuento de compras
 - Precio medio de compra
 - Valor total de las compras
 - Promedio de compras totales por persona por grupo de edad
6. Los que más gastan. Identifique a los 5 que más gastan en el juego por el valor total de la compra, y luego haga una lista (en una tabla):
 - SN
 - Recuento de compras

- Precio medio de la compra
- Valor total de la compra

7. Artículos más populares. Identifique los 5 artículos más populares según el recuento de compras y, a continuación, haga una lista (en una tabla):

- ID del artículo
- Nombre del artículo
- Número de compras
- Precio del artículo
- Valor total de la compra

8. Artículos más rentables. Identifique los 5 artículos más rentables por el valor total de la compra y, a continuación, haga una lista (en una tabla):

- ID del artículo
- Nombre del artículo
- Número de compras
- Precio del artículo
- Valor total de la compra

Como consideraciones finales:

1. Debes utilizar la librería Pandas y el Jupyter Notebook.
2. Debes enviar un enlace a tu Jupyter Notebook con los Data Frames visibles.
3. Debes incluir una descripción escrita de tres tendencias observables basadas en los datos.

3 Reto: El cambio climático, en datos y gráficos.

En la página: <https://www.epdata.es/datos/cambio-climatico-datos-graficos/447?accion=2>, se recoge en datos y gráficos algunas de las tendencias sobre las que vienen advirtiendo organismos internacionales y estudios científicos en relación a temas relacionados con el cambio climático como el aumento de la temperatura global, las emisiones de gases de efecto invernadero, los desplazamientos relacionados con eventos climáticos extremos o el aumento del nivel del mar.

Lo que queremos es ir a esa página y reproducir algunos de los gráficos y cálculos que aparecen allí usando pandas.