Logotipo, nombre de la empresa

Descripción generada automáticamente**PROGRAMACIÓN PARALELA – MEMORIA PRÁCTICA BICIMAD Y SPARK**

Esteban Joaquín Jiménez Párraga

Carlos Adolfo Gallego Gallego

Daniel Carretero Álvarez

**Problema:**

Una empresa de bicicletas desea publicitar sus nuevos cascos de ciclismo, recién salidos al mercado, ubicando vallas publicitarias por la ciudad de Madrid durante los meses de enero, febrero y marzo. Con el objetivo de maximizar el beneficio económico, la empresa ha decidido situar los anuncios en las rutas más transitadas por ciclistas madrileños, y así hacerse notar de forma eficiente ante potenciales nuevos clientes.

Casco de color verde

Descripción generada automáticamente con confianza mediaPara determinar cuáles son estas rutas, la empresa pretende analizar los datos proporcionados por el ayuntamiento de Madrid sobre el uso de las bicicletas de préstamo BICIMAD durante los tres primeros meses del año 2018; esta información incluye desde datos relativos al movimiento de la bicis, como la estación de partida y la de destino, hasta datos relativos al usuario que realiza el movimiento, como el tiempo de uso y la franja horaria en la que se ha tomado la bicicleta.

**Método:**

La idea es la siguiente: primero descargamos los tres ficheros con los datos correspondientes de cada mes. Trabajando con la librería *pyspark,* seleccionamos solamente los datos que nos interesan de cada uno de los ficheros para obtener la ruta óptima: el usuario, las estaciones de salida y destino, y las veces que se utilizan las bicicletas de estas. Uniendo la información de los tres meses en un único rdd mediante un *groupByKey()*, y calculamos las medias de uso mediante la función *mapValues()* aplicado sobre el método *mean*. Finalmente, empleamos *sortby()* para ordenarlas de mayor a menor según esta media y consultamos los 10 primeros elementos de nuestro rdd final aprovechando la evaluación perezosa de resultados.

**Resultados:**

Tras nuestro análisis mediante la ejecución de nuestro programa observamos que la ruta que en media más se ha usado durante los meses de Enero, Febrero y Marzo corresponde a la que comienza en la estación con código 9 y finaliza en la 149 con una media de 147,66 usos. Por ello recomendamos a la empresa la inversión en este recorrido como opción optima en cuanto a exposición.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Top 10 Rutas más recorridas y media de uso en los mese de Enero, Febrero y Marzo | | | | |
| **RUTAS** | **Inicio** | **Fin** | **Clave** | **Media** |
| **Ruta 1** | **9** | **149** | **(9,149)** | **147,66** |
| **Ruta 2** | **160** | **163** | **(160,163)** | **123,33** |
| **Ruta 3** | **175** | **43** | **(175,43)** | **122,33** |
| **Ruta 4** | **163** | **160** | **(163,160)** | **121,66** |
| **Ruta 5** | **130** | **149** | **(130,149)** | **121** |
| **Ruta 6** | **57** | **43** | **(57,43)** | **118,66** |
| **Ruta 7** | **135** | **132** | **(135,132)** | **117,33** |
| **Ruta 8** | **58** | **58** | **(58,58)** | **108,33** |
| **Ruta 9** | **135** | **135** | **(135,135)** | **108** |
| **Ruta 10** | **175** | **135** | **(175,135)** | **105,33** |