

Examen DS Avanzado

Instrucciones:

- Resuelve las preguntas de las secciones A, B y C.
- Envía tus respuestas a más tardar 48 horas después de recibir el correo.
- Las respuestas podrán ser enviadas en un PDF, un notebook, o con un link a tu repositorio de GitHub.
- Debes especificar claramente las respuestas de cada inciso, e indicar el código usado.

Sección A

A.1 Bebés

1. Para cada AGEB del Municipio de Toluca, estima cuántos bebés de 0 a 6 meses de edad habitan ahí el día de hoy. Explica tu razonamiento en menos de 300 palabras. Enlista tus fuentes y presenta los resultados. (Hint: revisa el [CPV 2010](#) , y el [Inventario Nacional de Vivienda 2016](#)). Puedes usar cualquier otra fuente que consideres relevantes.

Sección B

B1. QQP

Descarga la Base de datos histórica de [Quién es Quién en los Precios](#) de Profeco y resuelve los siguientes incisos. Para el procesamiento de los datos y el análisis exploratorio debes usar Spark SQL en el lenguaje de programación de tu elección.

1. Procesamiento de los datos
 - a. ¿Cuántos registros hay?
 - b. ¿Cuántas categorías?
 - c. ¿Cuántas cadenas comerciales están siendo monitoreadas?
 - d. ¿Cómo podrías determinar la calidad de los datos? ¿Detectaste algún tipo de inconsistencia o error en la fuente?
 - e. ¿Cuáles son los productos más monitoreados en cada entidad?
 - f. ¿Cuál es la cadena comercial con mayor variedad de productos monitoreados?
2. Análisis exploratorio
 - a. Genera una canasta de productos básicos que te permita comparar los precios geográfica y temporalmente. Justifica tu elección y procedimiento
 - b. ¿Cuál es la ciudad más cara del país? ¿Cuál es la más barata?
 - c. ¿Hay algún patrón estacional entre años?
 - d. ¿Cuál es el estado más caro y en qué mes?

- e. ¿Cuáles son los principales riesgos de hacer análisis de series de tiempo con estos datos?
- 3. Visualización
 - a. Genera un mapa que nos permita identificar la oferta de categorías en la zona metropolitana de León Guanajuato y el nivel de precios en cada una de ellas. Se darán puntos extra si el mapa es interactivo

Sección C

C.1 BOPS

Resuelve el siguiente [caso](#):

1. ¿Deberían expandirse a Canadá?
2. ¿Cuántos millones de dólares se ganaron o perdieron a partir del programa? Explica tu razonamiento y metodología.

Pista: Existen dos experimentos naturales.