Hito 1

**Abstracción de Tareas:**

1. Contrarrestar distintos factores socio económicos, étnicos y criminales de los distintos condados con muertes.

“Ver si existe una relación geográfica y factores socio económicos:”

Tarea: Descubrir tendencias entre los factores socio económicos y las zonas geográficas donde fallecidos personas a manos de la policía.

Justificación: Dado que no hay conocimiento cierto sobre la posible relación entre los factores socioeconómicos y las zonas geográficas, hace sentido que la acción correspondiente sea descubrir. Además, no podría ser comparar dado que solo encontramos un único objetivo. Por otra parte, se opto por tendencias como objetivo, dado que no sirve analizar simplemente si una zona geográfica presenta ciertos factores socio económicos, si no que para que un descubrimiento sea valido esta observación debe ser persistente en un numero suficiente de zonas geográficas y por ende se requiere una tendencia.

Datos: City, State, county\_fp, p\_income, county\_income, nat\_bucket, pov, urate, college

\*Se assume que podría desearse la información de los diferentes niveles de zonas geográficas

1. Relación entre número de muertes de minorías raciales y personas blancas

“… le interesan las características de los difuntos: específicamente le interesa ver si el número de personas difuntas pertenecientes a alguna minoría racial (hispánica, negra o asiática) es significativamente mayor al de personas blancas”

“le gustaría saber si dentro de cada etnicidad, algún sexo biológico es particularmente afectado en el número de muertes por sobre otro.”

Tarea: Comparar el valor individual de la cantidad de difuntos pertenecientes a cada raza y la distribución del sexo biológico dentro de cada uno de estos valores.

Justificación: Se selecciono la acción de comparar dado que hay mas de un objetivo. El primer objetivo corresponde a identificar si el numero de difuntos pertenecientes a alguna minoría racial es significativamente mayor que el numero de difuntos blancos. Es segundo lugar, se identifico el objetivo de poder identificar si la distribución de los sexos biológicos de los difuntos de las diferentes etnicidades no es similar. En conjunto con lo anterior se utilizo el objetivo el valor individual de los fallecimientos de personas de una sola etnia dado que solo interesa la cantidad por etnia y no su variación. Es el segundo objetivo se utilizo la distribución, dado que es interesante notar del total mencionado anteriormente como se distribuyen los sexos biológicos de los fallecidos.

Datos: raceethnicity, gender

1. Relación Fechas y concentración de incidentes

“piensa que podría haber un patrón o concentración de incidentes en fechas claves, como efemérides del país u otras celebraciones específicas”

Tarea: Descubrir (valor individual de) cantidad de incidentes a lo largo del año.

Justificación: Dado que tan solo se piensa y no hay un conocimiento claro respecto a esta creencia, la acción correspondiente debe ser descubrir. Por otra parte, dado que se agrupan la cantidad de incidentes por fecha, no se busca encontrar una tendencia en los datos si no que tan solo poder ver cuantos incidentes hubo por fecha (suma de incidentes por fecha) y con esa información notar si existe una mayor concentración en ciertas fechas específicas.

Datos: suma dates agrupadas por fecha

**Abstracción de Datos:**

* Name
  + Categórico: Los nombres no tienen un orden intrínseco y numérico
  + El resto de las clasificaciones no aplica a este dato
* Age
  + Ordenado: Al ser una secuencia de números, tiene un orden intrínseco y numérico
  + Cuantitativos: Existe comparación aritmética exacta entre las edades
  + Secuencial: Dado que tiene un mínimo y crece en una sola dirección
  + No cíclico: Ya que teóricamente podría crecer indefinidamente y no se vuelve al origen.
* Gender
  + Categórico: Los sexos biológicos no tienen un orden intrínseco ni matemático
  + El resto de las clasificaciones no aplican para este dato
* Raceethnicity
  + Categórico: No presentan un orden intrínseco
  + El resto de las categorías no aplican
* Date
  + Ordenado: dado que las fechas (días y mes) tienen un orden intrínseco (desde el comienzo a final de año).
  + Ordinales: No existe una comparación aritmética exacta entre las fechas
  + Secuencial: no existe un cero que divida el conjunto de fechas, ni un crecimiento en ambas direcciones
  + Cíclico: Al ser solo día y mes, al terminar un año se vuelve a comenzar desde la fecha mínima (01/01)
* City
  + Categórico: la ciudad no cuenta con un orden intrínseco y numérico
  + El resto de las clasificaciones no aplican
* State
  + Categórico: Los estados no cuentan con un orden intrínseco ni matemático
  + El resto de las clasificaciones no aplican
* Latitude
  + Ordenado: Dado que la latitud, rango de números reales, tienen un orden intrínseco y matemático
  + Cuantitativos: al ser números, si existe una comparación aritmética exacta
  + Divergente: dado que están las latitudes positivas y las negativas, pero ambas comparten la línea del ecuador (latitud 0) como cero en común
  + No cíclico: Dado que no existe un lugar de inicio o valor de latitud inicial, se podría considerar un polo o el ecuador, pero de todas formas pese a estar acotado no existe un ciclo.
* Longitud
  + Ordenado: Dado que la longitud, rango de números reales, tienen un orden intrínseco y matemático
  + Cuantitativos: al ser números, si existe una comparación aritmética exacta
  + Divergente: dado que existen longitudes positivas y las negativas, pero ambas comparten el meridiano de Greenwich (longitud 0) como cero en común
  + No cíclico: Dado que no existe un lugar de inicio o valor de longitud inicial, se podría considerar el meridiano o cualquier otra longitud, pero de todas formas pese a estar acotado no existe un ciclo.
* County\_fp
  + Categórico: al ser un código identificador, este no tiene un orden intrínseco
  + El resto de las clasificaciones no aplican
* Cause
  + Categórico: Una frase no tiene un orden intrínseco ni matemático
  + El resto de las categorías no aplican
* Armed
  + Categórico: Un booleano o tipo de arma no tienen un orden intrínseco
  + El resto de las categorías no aplican
* Share\_white
  + Ordenado: Los porcentajes al ser números tienen un orden intrínseco y matemático
  + Cuantitativo: Los porcentajes permiten comparaciones aritméticas exactas
  + Secuencial: Los porcentajes van de 0 a 1 (o 0 a 100), pero no existen dos direcciones de crecimiento ni un cero que separe este conjunto
  + No cíclicos: Ya que al alcanzar el 100%, simplemente no se puede continuar aumentando, pero no se vuelve al 0%
* Share\_black
  + Ordenado: Los porcentajes al ser números tienen un orden intrínseco y matemático
  + Cuantitativo: Los porcentajes permiten comparaciones aritméticas exactas
  + Secuencial: Los porcentajes van de 0 a 1 (o 0 a 100), pero no existen dos direcciones de crecimiento ni un cero que separe este conjunto
  + No cíclicos: Ya que al alcanzar el 100%, simplemente no se puede continuar aumentando, pero no se vuelve al 0%
* Share\_hispanic
  + Ordenado: Los porcentajes al ser números tienen un orden intrínseco y matemático
  + Cuantitativo: Los porcentajes permiten comparaciones aritméticas exactas
  + Secuencial: Los porcentajes van de 0 a 1 (o 0 a 100), pero no existen dos direcciones de crecimiento ni un cero que separe este conjunto
  + No cíclicos: Ya que al alcanzar el 100%, simplemente no se puede continuar aumentando, pero no se vuelve al 0%
* P\_income
  + Ordenado: El ingreso es un numero natural el cual tiene un orden intrínseco y matemático, por lo que la mediana de este también tiene
  + Cuantitativo: El ingreso permite comparaciones aritméticas exactas, por lo que la mediana sigue siendo un ingreso y por ende también permite comparaciones aritméticas exactas.
  + Secuencial: No existe un 0 que divida el conjunto en dos direcciones de crecimiento.
  + No cíclico: La cantidad del ingreso, y su mediana, puede crecer indefinidamente
* County\_income
  + Ordenado: El ingreso es un numero natural el cual tiene un orden intrínseco y matemático, por lo que la mediana de este también tiene
  + Cuantitativo: El ingreso permite comparaciones aritméticas exactas, por lo que la mediana sigue siendo un ingreso y por ende también permite comparaciones aritméticas exactas.
  + Secuencial: No existe un 0 que divida el conjunto en dos direcciones de crecimiento.
  + No cíclico: La cantidad del ingreso, y su mediana, puede crecer indefinidamente
* Nat\_bucket
  + Ordenado: los quintiles representan ciertos porcentajes, y por lo tanto tienen un orden intrínseco y matemático
  + Cuantitativo: Dado que el valor de este atributo es un numero representando el quintil a nivel país, los números utilizados en esta representación permiten comparaciones aritméticas exactas
  + Secuencial: No existe un cero que permite dividir los posibles quintiles
  + No cíclico:
* Pov
  + Ordenado: Los porcentajes al ser números tienen un orden intrínseco y matemático
  + Cuantitativo: Los porcentajes permiten comparaciones aritméticas exactas
  + Secuencial: Los porcentajes van de 0 a 1 (o 0 a 100), pero no existen dos direcciones de crecimiento ni un cero que separe este conjunto
  + No cíclicos:
* Urate
  + Ordenado: Las razones al ser números tienen un orden intrínseco y matemático
  + Cuantitativo: Las razones permiten comparaciones aritméticas exactas
  + Secuencial: Las razones van de 0 a 1, pero no existen dos direcciones de crecimiento ni un cero que separe este conjunto
  + No cíclicos: No existe un valor que, al ser alcanzado, cause que se vuelva al valor mínimo o inicial
* College
  + Ordenado: Las razones al ser números tienen un orden intrínseco y matemático
  + Cuantitativo: Las razones permiten comparaciones aritméticas exactas
  + Secuencial: Las razones van de 0 a 1, pero no existen dos direcciones de crecimiento ni un cero que separe este conjunto
  + No cíclicos: No existe un valor que, al ser alcanzado, cause que se vuelva al valor mínimo o inicial

¿De qué tipo de dataset se trata pólice\_killings.csv?

El tipo de dataset corresponde a una tabla, dado que como se puede notar esta en un archivo csv. Este se caracteriza por almacenar información en forma de tablas mediante filas y columnas.

¿Es un dataset estático o dinámico?

El dataset es estático dado que contamos con los datos antes de comenzar a trabajar, y no tenemos información de que estos se vayan a actualizar en el futuro.

¿Qué tipos de datos singulares hay presentes en el dataset?

En el dataset encontramos datos del tipo item, correspondiente a cada una de las filas en su totalidad dado que cada caso es una entidad compleja que esta compuesta de múltiples valores. También podemos encontrar atributos, correspondientes a las celdas en la tabla, dado que cada una es una medición o valor en específico. También se cuenta con datos del tipo posición, dado que contamos con latitud y longitud, ambos definen una posición en una dimensión del espacio y uniendo ambas, se obtiene una posición en especifico dentro de los estados unidos.