Proyecto en grupo 2: Métodos analíticos de optimización y simulación

Camilo Ramírez Guilombo

Farid Alonso Saad Plata

Maria Gabriela Martínez Lopera

Profesora Yris Olaya

Profesora Gloria Patricia Jaramillo

Universidad Nacional de Colombia

Facultad de Minas

Departamento de Ciencias de la Computación y la Decisión

Medellín

Septiembre de 2016

Tabla de contenido

[1. Descripción del problema 3](#_Toc463388110)

[2. Análisis descriptivo de los datos 4](#_Toc463388111)

[2.1. Información para el viernes 5 de enero 8](#_Toc463388112)

[2.2. Información para el sábado 6 de enero 9](#_Toc463388113)

[2.3. Información para el domingo 7 de enero 11](#_Toc463388114)

[3. Planteamiento de solución y recomendaciones 12](#_Toc463388115)

[4. Modelo conceptual de simulación 13](#_Toc463388116)

[4.1. Validación del modelo de simulación 14](#_Toc463388117)

[4.1.1. Validación del usuario 14](#_Toc463388118)

[4.1.2. Validación en el modelo conceptual 14](#_Toc463388119)

[4.1.3. Validación operacional 15](#_Toc463388120)

[5. Formulación de un problema de optimización 15](#_Toc463388121)

[5.1. Variable de decisión 16](#_Toc463388122)

[5.2. Parámetros 16](#_Toc463388123)

[5.3. Función objetivo 16](#_Toc463388124)

[5.4. Restricciones 16](#_Toc463388125)

# Descripción del problema

Se tomará como insumo, la información contenida en el archivo UberTripDistrictData\_SnFrancisco.csv, que incluye el detalle de 24682 servicios de transporte prestados por la compañía Uber, en el área metropolitana de San Francisco, CA, para el periodo de tiempo comprendido entre el 1 y el 7 de enero del año 2007.

El propósito principal de este análisis será determinar, con base en el análisis descriptivo de la información presentada, las franjas horarias y los sectores en los cuales se deben ubicar los conductores de Uber, con el fin de que puedan maximizar sus viajes o servicios durante un lapso específico.  Este objetivo está limitado por dos restricciones clave:

* Sólo se estudiará un tipo de servicio prestado por la compañía (aquel que estadísticamente presente más registros en el archivo que se analizará).
* Únicamente se harán recomendaciones de asignaciones para las madrugadas de los días viernes 5, sábado 6 y domingo 7 de enero, que como se verá más adelante, albergan la mayor cantidad de servicios prestados.

Se debe definir primero cuál es tipo de servicio de Uber que se estudiará. A continuación se muestran los registros existentes para cada tipo de servicio en el archivo inicial.

Tabla 1. Número de viajes registrados para cada tipo de servicio prestado por la compañía Uber. Fuente: elaboración propia en R, a partir del archivo UberTripDistrictData\_SnFrancisco.csv.

|  |  |
| --- | --- |
| *Tipo de servicio* | *Número de registros* |
| UberX | 14672 |
| UberSelect | 2895 |
| UberXL | 4610 |
| UberSUV | 1731 |
| UberBlack | 774 |

De acuerdo con la información presentada en la Tabla 1, se realizará el análisis para el tipo de servicio UberX.

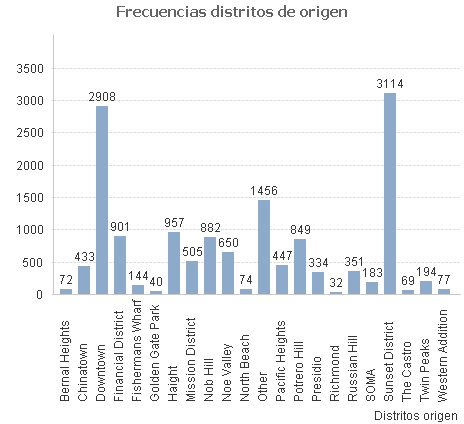
Se cuenta con los siguientes datos:

* Duración y distancia del viaje
* Distrito de salida y distrito de llegada.
* La fecha en la cual se dio el servicio.
* La hora en la cual se recoge el pasajero y la hora en la que se descarga.

# Análisis descriptivo de los datos

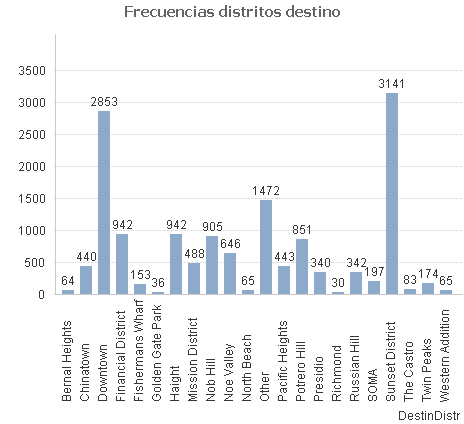
Se comenzará por determinar cuáles son los distritos desde los cuales se generan más servicios de UberX, y en cuales hay mayor número de servicios finalizados. Para esto se hace un diagrama de frecuencias de los servicios  para los distritos origen y destino.

Gráfica 1. Histograma de frecuencias para los distritos origen. Fuente: elaboración propia en QlikView, a partir del archivo UberTripDistrictData\_SnFrancisco.csv.



A partir del anterior histograma, es posible concluir que en “Sunset district” y “Downtown”, los conductores de Uber recogieron al mayor número de personas para el intervalo de tiempo estudiado.

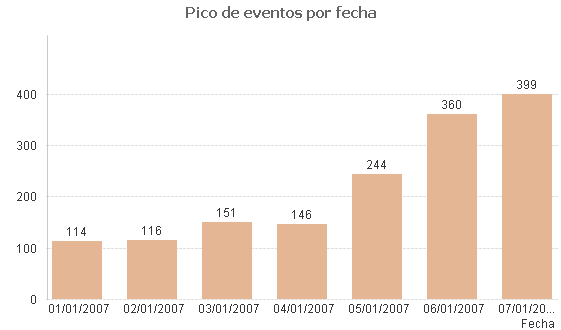
Gráfica 2. Histograma de frecuencias para los distritos destino. Fuente: elaboraciónpropia en QlikView, a partir del archivo UberTripDistrictData\_SnFrancisco.csv.

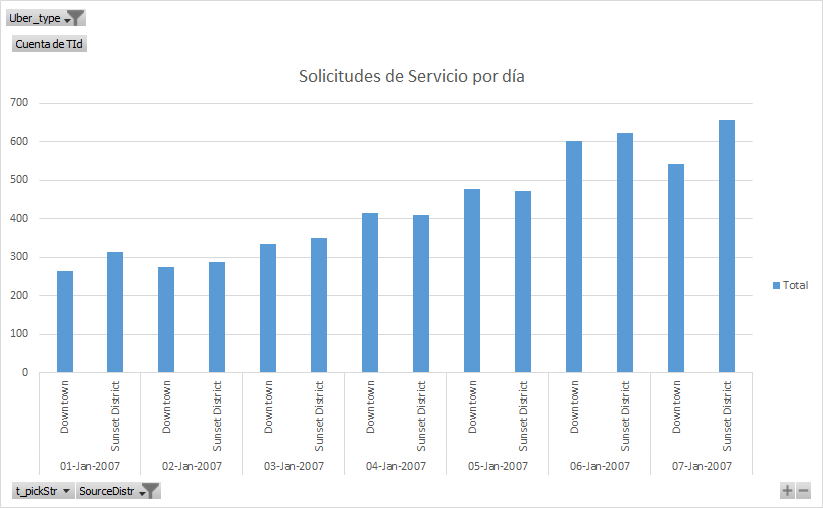


Igualmente, el anterior histograma de frecuencias de destinos, confirma que los destinos finales más comunes son los distritos “Sunset District” y “Downtown”.

Al tener esto, el análisis se centrará en las interacciones entre estos dos distritos, que tienen la mayor frecuencia de servicios.

Gráfica 3. Picos de servicios registrados por hora para todos los distritos. Fuente: elaboración propia en QlikView, a partir del archivo UberTripDistrictData\_SnFrancisco.csv.



Gráfica 4. Solicitudes del servicio de Uber, por día y para los distritos más visitados. Fuente: elaboración propia en Microsoft Excel, a partir del archivo UberTripDistrictData\_SnFrancisco.csv.

Gráfica 5. Picos de servicios registrados por hora del día para los fines de semana y todos los distritos. Fuente: elaboración propia en QlikView, a partir del archivo UberTripDistrictData\_SnFrancisco.csv.



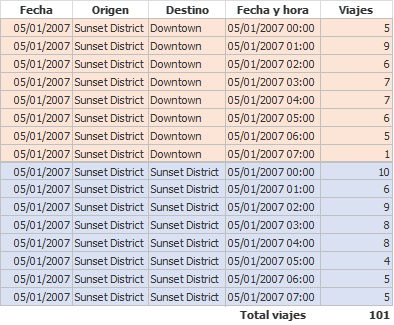
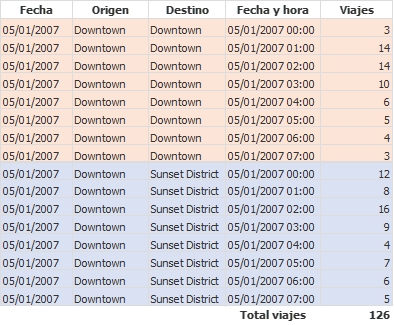
La información registrada en las gráficas 3, 4 y 5, explica por qué se limitó el análisis del sistema estudiado a los días viernes, sábado y domingo en horas de la madrugada.

En general, los días del fin de semana resultan ser los que mayor demanda de servicio tienen, en términos de todos los distritos (Gráfica 3) y también de los de interés (Sunset y Downtown) (Gráfica 4). Además, resulta claro que el servicio de UberX es más demandado en horas de la madrugada, que en el resto del día, por lo que se recomendarán patrones de conducta para la franja horaria comprendida entre la medianoche y las 7 de la mañana de los días viernes, sábado y domingo.

Con el objetivo de encontrar el distrito que maximice el número de viajes para un conductor, en cada franja horaria considerada para el fin de semana, se muestran a continuación los diferentes viajes realizados entre los distritos Sunset y Downtown, y dentro de cada uno de ellos.

* 1. Información para el viernes 5 de enero

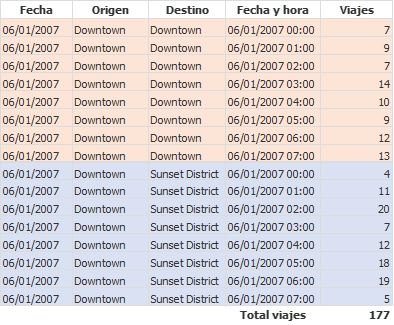
Tabla 2. Detalle del número de viajes realizados entre y dentro los distritos Sunset y Downtown para el día viernes 5 de enero de 2007 en horas de la madrugada. Fuente: elaboración propia en QlikView, a partir del archivo UberTripDistrictData\_SnFrancisco.csv.

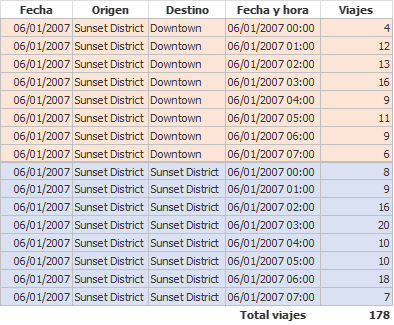


De acuerdo con la información presentada en la Tabla 2, existe una diferencia significativa de 25 viajes entre el total de servicios prestados por un conductor de Uber, cuando el origen es el Downtown y el Sunset District. Este hecho podría sugerir que la localización que posibilita la maximización del número de viajes en la madrugada del viernes 5 de enero, es el Downtown.

* 1. Información para el sábado 6 de enero

Tabla 3. Detalle del número de viajes realizados entre y dentro los distritos Sunset y Downtown para el día sábado 6 de enero de 2007 en horas de la madrugada. Fuente: elaboración propia en QlikView, a partir del archivo UberTripDistrictData\_SnFrancisco.csv.

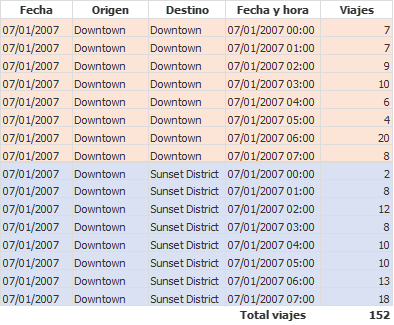


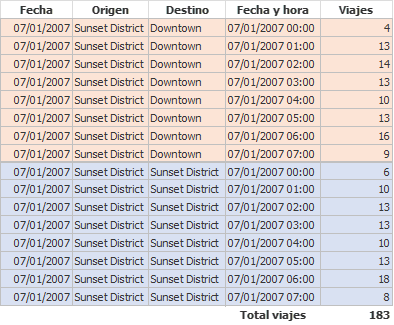


De acuerdo con la información presentada en la Tabla 3, se puede concluir que en la madrugada del día sábado 6 de enero, la cantidad de viajes que parten desde Sunset y Downtown no difieren significativamente, por lo que se plantea que es indiferente en cuál de los dos se localice el conductor, pues en ambos distritos encontrará una demanda muy similar de servicios. Además una continua movilidad entre los dos distritos asegurará viajes, ya que el sábado es el día con mayor demanda del fin de semana.

* 1. Información para el domingo 7 de enero

Tabla 4. Detalle del número de viajes realizados entre y dentro los distritos Sunset y Downtown para el día domingo 7 de enero de 2007 en horas de la madrugada. Fuente: elaboración propia en QlikView, a partir del archivo UberTripDistrictData\_SnFrancisco.csv.





De acuerdo con la información presentada en la Tabla 4, puede concluirse que el distrito Sunset presenta una considerable diferencia en servicios solicitados respecto al distrito Downtown, por lo que se recomienda que el conductor permanezca en el distrito Sunset durante la madrugada del domingo 7 de enero.

# Planteamiento de solución y recomendaciones

Cabe aclarar que el análisis de datos aquí presentado se basó mayormente en la cantidad de servicios prestados, y no se utilizaron variables que se tenían en la base de datos tales como la distancia recorrida y el tiempo de viaje. La eventual inclusión de estas variables al análisis podría permitir recomendaciones más realistas y específicas sobre las rutas y horarios que le generarían mayores ingresos a un conductor de Uber, pues la maximización de los viajes no necesariamente representa una maximización de las ganancias, dado que con esta última se relacionan estrechamente las variables de duración y el recorrido total de cada servicio. Esta inclusión se propone como una recomendación a estudios más detallados.

Otra consideración importante tiene que ver con el hecho de que no se pueden generalizar las recomendaciones de estas ubicaciones para otras semanas del año. En primer lugar, porque no se tienen datos suficientes para simular el comportamiento de la demanda en un lapso significativo, y en segundo lugar, porque la primera semana del año suele comportarse diferente respecto a las demás, pues se trata del inicio del año. Esto implica que un análisis predictivo sólo podría tomar lugar cuando se tenga una muestra significativa de datos relacionados con la demanda de este servicio de Uber, para que de esta manera pueda simularse el comportamiento del sistema en el tiempo. El alcance del presente informe no puede ir más allá de la exploración de los datos estudiados.

En resumidas cuentas, el análisis descriptivo llevado a cabo, permite concluir las siguientes recomendaciones para un conductor de Uber que quiera maximizar el número de servicios prestados durante el fin de semana del 5 al 7 de enero de 2007, en las franjas horarias que van desde la media noche hasta las 7 am:

Tabla 5. Ubicaciones ideales encontradas para el fin de semana entre el 5 y el 7 de Enero de 2007. Fuente: elaboración propia en QlikView, a partir del archivo UberTripDistrictData\_SnFrancisco.csv.

|  |  |
| --- | --- |
| **Día** | **Ubicación** |
| Viernes 5 | Downtown |
| Sábado 6 | Downtown o Sunset District |
| Domingo 7 | Sunset District |

# Modelo conceptual de simulación

Con los datos que se tienen sobre los viajes realizados por el sistema Uber en los diferentes distritos de San Francisco, se pueden realizar diferentes simulaciones dependiendo del objetivo que se plantee el decisor.

En este caso particular, se trató de maximizar una función objetivo que sumaba los viajes realizados desde un distrito a otro. Dado este escenario, se propondrá un modelo de simulación que permita a un conductor de Uber elegir el distrito en el cual deberá iniciar su día de trabajo en las madrugadas (12 a 7 a.m.) de los fines de semana basado en el análisis de datos anterior, en el cual se puede evidenciar que la demanda para esta franja de tiempo tiene un comportamiento superior al resto de la semana.

Para llevar a cabo esto, se requiere ajustar distribuciones de probabilidad y determinar los siguientes datos de entrada:

* Tiempo entre solicitudes del servicio en cada uno de los distritos.
* Probabilidad de que el usuario se dirija a cada uno de los distritos dado el lugar en el que se realizó la solicitud.
* Duración del tiempo de servicio entre distritos.
* Distancia recorrida entre distritos.

Los datos de salida o medidas de desempeño del modelo son:

* Ingresos totales.
* Porcentaje de ocupación.
* Porcentaje de servicios con destino a cada una de las ciudades.
* Servicios que no pudieron ser atendidos por el conductor

La construcción del modelo se debe enfocar en representar el siguiente sistema:

Se define el distrito de inicio deseado y se genera una solicitud de servicio por parte de un cliente, una vez se da el ingreso del primer cliente al sistema se decide el destino del usuario utilizando la probabilidad que se dirija a cada distrito, para así generar un tiempo de servicio y distancia recorrida desde la zona inicial a la zona final (un servicio puede presentarse internamente en el mismo distrito). Con estos datos es posible determinar el valor del servicio, el cual se deberá acumular a lo largo de la simulación, de igual modo se captura el tiempo de servicio para obtener al final el porcentaje de ocupación. Una vez se tiene libre el vehículo (se termina el servicio anterior), se inicia el mismo procedimiento, siempre el distrito origen será el distrito final del servicio anterior. Se deberá almacenar el distrito destino de cada usuario que ingresa al sistema, con el fin de determinar el porcentaje de servicios que tuvieron destino a cada ciudad.

La simulación deberá realizarse entre las 12 a.m y las 7 a.m de un día, es conveniente iterar un número considerable de veces este proceso con el fin de obtener resultados que converjan a un valor estable. Se deberá realizar la simulación iniciando en cada uno de los distritos, con el fin de comparar los resultados cuando se inicia la operación de cada uno de ellos.

La simulación tiene como supuestos que el conductor no se negará a prestar ningún servicio y que éste está dispuesto a ir a todos los distritos de San Francisco. La medida de desempeño orientada a establecer el porcentaje de servicios que se hace a cada una de las ciudades, permitirá tomar decisiones que estén por fuera del desempeño económico, es decir si el conductor tiene algún tipo de preferencia por estar la mayor parte del tiempo en un distrito particular, o si quiere reducir el porcentaje de tiempo que está en alguno de los distritos, ya sea por seguridad, lejanía a su hogar u otros factores.

## Validación del modelo de simulación

El planteamiento y determinación de todos los componentes de un modelo conceptual de simulación, como la identificación del problema, el establecimiento de objetivos, la determinación de entradas, salidas, restricciones, suposiciones y alcance del modelo, debe ir complementada por la validación, y este proceso se considera parte vital del desarrollo del proyecto. Dependiendo de los requerimientos de confiabilidad que se quieran en la validación, existirán variaciones en la cantidad de métodos aplicados, en el tiempo que toman y en su costo.

Existen diversos métodos de validación, a continuación se plantean algunas propuestas para el caso específico de Uber según el modelo expuesto anteriormente.

* + 1. Validación del usuario

Se involucran los conductores de Uber para contrastar los resultados dependiendo del lugar donde empiezan el servicio. Una de las medidas más certeras sería confirmar que el mejor lugar para empezar según los usuarios, sea el mejor según la simulación; esta misma verificación con los que no generan tantas ganancias ni tanto porcentaje de ocupación.

* + 1. Validación en el modelo conceptual

Se debe verificar que las relaciones matemáticas de las variables a utilizar son razonables a los propósitos del modelo. Para esto se debe verificar que el ajuste de los datos a distribuciones de probabilidad sean de alta confiabilidad y que por ejemplo el tiempo entre llegada de los servicios se distribuya de manera exponencial. Otra validación realizada al desarrollo del modelo conceptual se da por medio de la trazabilidad a las entidades, en este caso se puede hacer un seguimiento a los conductores de Uber y verificar que los diferentes caminos que este sigue en la realidad, se acercan significativamente a los caminos que sigue la entidad en el modelo conceptual.

* + 1. Validación operacional

Se quiere verificar si las respuestas del modelo tienen concordancia con los requerimientos del modelo planteado conceptualmente. Para esto hay numerosas técnicas, mayormente análisis de sensibilidad relacionados con los parámetros de las variables y pruebas en condiciones extremas.

* Condiciones extremas en la generación de servicios: Se pueden hacer modificaciones en el tiempo entre la solicitud de servicios, estableciendo que se ajustan bajo una distribución de probabilidad promedio con media muy baja, por ejemplo 0.01 minutos, es decir que cada 0.06 segundos se solicitará un servicio. El modelo será válido si los respuestas arrojan resultados significativamente altos en el porcentaje de ocupación y en el número de servicios no atendidos.
* Condiciones extremas en el tiempo de servicio: Se puede plantear que el tiempo de servicio entre los distritos sigue una distribución uniforme entre 100 y 1000 minutos, por lo que se tendrían tiempos de servicios muy altos. La validez bajo esta condición se evidenciaría en un alto porcentaje de ocupación para el conductor, pero también debe presentarse en que el porcentaje de servicios con destino a cada uno de los distritos sería muy bajo, y para alguna de estas sería nulo.
* Condiciones extremas en términos de los distritos: En términos de los distritos una manera muy sencilla de validar sería definir la probabilidad de dirigirse a un número determinado de distritos como nula, simular y verificar que en efecto al conductor de Uber nunca se dirigió a estos distritos.
* Análisis de sensibilidad: estas pueden ser variaciones porcentuales a los diferentes parámetros estimados para las variables involucradas en el modelo. La validación desde este punto se manejaría como la del siguiente ejemplo, por ejemplo si se aumenta en un determinado porcentaje el tiempo medio de entrada de servicios en el sistema, el porcentaje de ocupación del conductor disminuiría y el número de servicios sin atender disminuiría; otro ejemplo es si se aumenta en un determinado porcentaje el tiempo de servicio entre los distritos, habrían más servicios sin atender.

# Formulación de un problema de optimización

El objetivo de la optimización es encontrar el recorrido óptimo, es decir, determinar cuál es la ruta óptima que debe seguir el conductor de Uber para maximizar sus ganancias al final de la madrugada del día domingo. Bajo este esquema se asume que el conductor de Uber X podrá elegir la carrera a tomar de acuerdo al destino del pasajero.

* 1. Variable de decisión
  2. Parámetros

Corresponde al peso de demanda que tiene cada ruta .

Se calcula como:

, donde:

* 1. Función objetivo

**Maximizar las ganancias obtenidas al final del día sábado:**

* 1. Restricciones

**El tiempo de recorrido no debe ser superior a 420 minutos - 7 horas (12 a.m a 7 a-m)**

**Ecuaciones de balance**

**Una sola ruta**

El objetivo es que la solución sea una ruta cerrada, la cual no necesariamente sumará 420 minutos, sin embargo el conductor tendrá claro el orden en que debe abordar las carreras (bajo el supuesto que el conductor elige el destino). La solución no tiene un número de ciudades a visitar, lo que implica que la solución podría ser incluso realizar carreras dentro de un mismo distrito (caso en el que decide ir del distrito origen i al destino j donde i=j).