# INTRODUCCIÓN

<Introducir el problema a resolver> [1-2 párrafos]

Mediante la introducción de las nuevas tecnologías en ámbitos donde tradicionalmente se ha gestionado la información en papel, se ha logrado que el volumen de la información y la velocidad con la que se accede a ellos se multiplique. No obstante, aumentar el volumen de datos captados y almacenados no es suficiente; también es necesario hacer una traducción de los datos para transformarlos en conocimiento, y presentarlos para una interpretación rápida y asequible por parte de las personas que los necesiten y no sean conocedoras de la tecnología.

Con esta premisa, este trabajo pretende coleccionar e interpretar los datos recogidos por el departamento de policía de la ciudad de Nueva York (referidos a partir de ahora como NYPD, del inglés *New York Police Department*)[[1]](#footnote-1). Se trata del cuerpo de policía más grande en Estados Unidos, que actualmente cuenta con 36 000 agentes y 18 000 empleados, cuya responsabilidad es la de proteger una cuidad con casi 9 millones de habitantes. Su labor ha sido uno de los principales responsables en el descenso de crímenes en la cuidad, convirtiéndola en una de las más seguras del país.

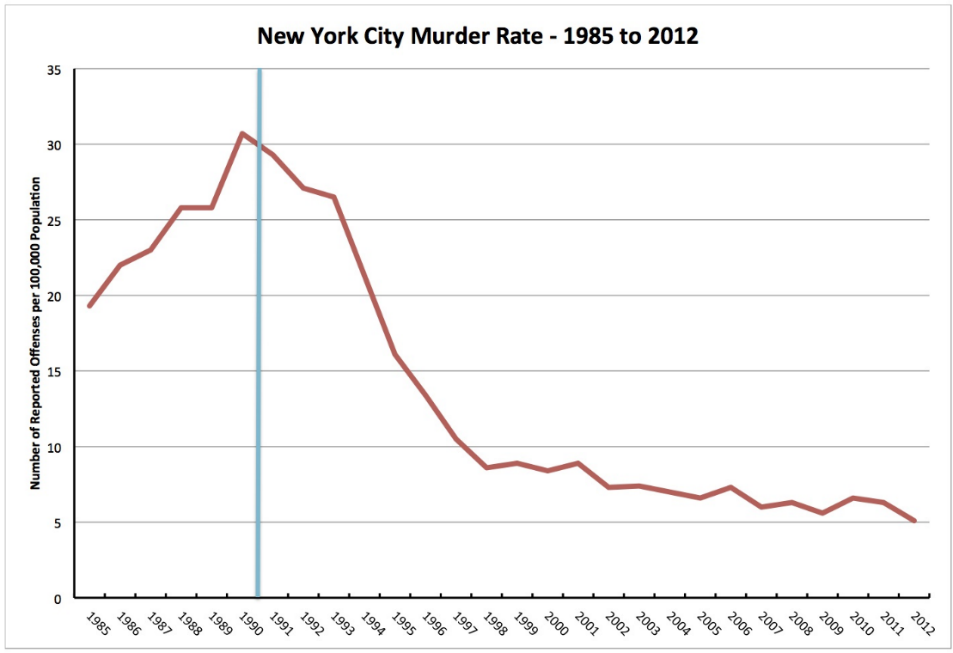


Ilustración 1: Tasa de homicidios en NYC 1985-2012[[2]](#footnote-2)

<Comentar los desafíos y las necesidades del problema que justifique el uso de una solución de inteligencia de negocios> [1 párrafo]

La labor del NYPD ha generado una considerable cantidad de datos, que se encuentran disponibles como *open data* en la página web de la ciudad[[3]](#footnote-3). Es necesario hacer una interpretación de estos datos para presentar al personal encargado de gestionar la seguridad de la urbe. Este proyecto se centrará en analizar los datos de los crímenes principalmente según los tipos de ofensas cometidas y su situación geográfica dentro de la ciudad, haciendo uso de los datos proporcionados en la web.

<Indicar cómo ha sido abordado el problema (alto nivel), describiendo los objetivos perseguidos y los tipos de análisis considerados> [1-2 párrafos]

<Explicar las tecnologías usadas en cada etapa de la solución de IN desarrollada> [1 párrafo]

# PLAN

## OBJETIVOS

<Indicar los objetivos generales y los objetivos específicos>

Los objetivos identificados para el trabajo se dividen en objetivos principales y objetivos secundarios. Estos se exponen a continuación:

### OBJETIVOS PRINCIPALES

|  |  |
| --- | --- |
| **OBJETIVO P-01** | **Optimizar la ubicación de las comisarías de policía** |
| Descripción | Se pretende optimizar la ubicación geográfica de las comisarías de policía en la ciudad en relación con los focos de delincuencia. |
| Autores | Grupo 0X |
| Versión | 1.0 |

* + 1. OBJETIVOS SECUNDARIOS

|  |  |
| --- | --- |
| **OBJETIVO S-01** | **Visualizar las zonas conflictivas y más seguras de la cuidad** |
| Descripción | Mediante un mapa de calor, se mostrarán las zonas con mas prevalencia de delitos, así como las más seguras. Este mapa deberá ser interactivo y accesible a través de una aplicación web con un navegador. |
| Autores | Grupo 0X |
| Versión | 1.1 |

|  |  |
| --- | --- |
| **OBJETIVO S-02** | **Visualizar los datos en tiempo real en un cuadro de mandos** |
| Descripción | Los datos se tomarán directamente de la web NYC Open Data y se procesarán de manera automática, para así poder mantenerlos actualizados. Además, estos se visualizarán en un cuadro de mandos interactivo que presente la información de manera accesible mediante gráficos desde un navegador. |
| Autores | Grupo 0X |
| Versión | 1.1 |

|  |  |
| --- | --- |
| **OBJETIVO S-03** | **Diseñar un modelo predictivo para detectar la ofensa cometida** |
| Descripción | Se intentará predecir el tipo de delito cometido para un registro mediante el diseño y desarrollo de un modelo predictivo y técnicas de selección de atributos. |
| Autores | Grupo 0X |
| Versión | 1.0 |

## PLAN DE TRABAJO

<Especificar cómo se lograrán alcanzar cada uno de los objetivos>

Para alcanzar cada todos los objetivos, se dividirá el proyecto en paquetes de trabajo según los objetivos, y estos se repartirán entre los miembros del equipo de trabajo. Los miembros que no participen en un objetivo tendrán la responsabilidad de dar su visto bueno al trabajo realizado por los compañeros.

<Dividir el proyectos en paquetes de trabajo y éstos en tareas>

Las tareas se han dividido de la siguiente forma:

|  |  |  |  |
| --- | --- | --- | --- |
| **Nombre de tarea** | **Duración** | **Comienzo** | **Fin** |
| **Lanzamiento** | **1 día** | **lun 18-04-16** | **lun 18-04-16** |
| Redacción de objetivos | 1 día | lun 18-04-16 | lun 18-04-16 |
| **Redacción de la documentación** | **6 días** | **jue 18-04-26** | **lun 18-05-21** |
| Plan de Proyecto | 1 día | jue 18-04-26 | jue 18-04-26 |
| Análisis | 1 día | vie 18-04-27 | vie 18-04-27 |
| Diseño | 2 días | lun 18-04-30 | mar 18-05-01 |
| Implementación | 1 día | vie 18-05-18 | vie 18-05-18 |
| Despliegue | 0.5 días | lun 18-05-21 | lun 18-05-21 |
| Conclusiones | 0.5 días | lun 18-05-21 | lun 18-05-21 |
| **Búsqueda de conjunto de datos** | **7 días** | **mar 18-04-17** | **mié 18-04-25** |
| Búsqueda y adquisición de datos | 1 día | mar 18-04-17 | mar 18-04-17 |
| Búsqueda de información de tecnologías | 1 día | mié 18-04-18 | mié 18-04-18 |
| Aprendizaje de tecnologías | 2 días | jue 18-04-19 | vie 18-04-20 |
| Preprocesado de los datos | 3 días | lun 18-04-23 | mié 18-04-25 |
| **Objetivo P-01** | **7 días** | **mié 18-05-09** | **jue 18-05-17** |
| Búsqueda de herramientas | 1 día | mié 18-05-09 | mié 18-05-09 |
| Aprendizaje de las herramientas | 2 días | jue 18-05-10 | vie 18-05-11 |
| Implementación de la solución | 3 días | lun 18-05-14 | mié 18-05-16 |
| Pruebas de la solución | 1 día | jue 18-05-17 | jue 18-05-17 |
| **Objetivo S-01** | **5 días** | **mié 18-05-02** | **mar 18-05-08** |
| Búsqueda de herramientas | 1 día | mié 18-05-02 | mié 18-05-02 |
| Aprendizaje de las herramientas | 1 día | jue 18-05-03 | jue 18-05-03 |
| Implementación de la solución | 2 días | vie 18-05-04 | lun 18-05-07 |
| Pruebas de la solución | 1 día | mar 18-05-08 | mar 18-05-08 |
| **Objetivo S-02** | **5 días** | **mié 18-05-02** | **mar 18-05-08** |
| Búsqueda de herramientas | 1 día | mié 18-05-02 | mié 18-05-02 |
| Aprendizaje de las herramientas | 1 día | jue 18-05-03 | jue 18-05-03 |
| Implementación de la solución | 2 días | vie 18-05-04 | lun 18-05-07 |
| Pruebas de la solución | 1 día | mar 18-05-08 | mar 18-05-08 |
| **Objetivo S-03** | **5 días** | **mié 18-05-02** | **mar 18-05-08** |
| Búsqueda de herramientas | 1 día | mié 18-05-02 | mié 18-05-02 |
| Aprendizaje de las herramientas | 1 día | jue 18-05-03 | jue 18-05-03 |
| Implementación de la solución | 2 días | vie 18-05-04 | lun 18-05-07 |
| Pruebas de la solución | 1 día | mar 18-05-08 | mar 18-05-08 |

Los paquetes de trabajo se han distribuido de la siguiente manera:

|  |  |  |
| --- | --- | --- |
| **Paquete de tareas** | **Duración** | **Responsable** |
| Lanzamiento | 1 día | Grupo 0X |
| Redacción de la documentación | 6 días | Grupo 0X |
| Búsqueda de conjunto de datos | 7 días | Grupo 0X |
| Objetivo P-01 | 7 días | Grupo 0X |
| Objetivo S-01 | 5 días | Juan Antonio Rodríguez |
| Objetivo S-02 | 5 días | Alberto Cárdenas |
| Objetivo S-03 | 5 días | Manuel Ridao |

<Indicar la planificación temporal>

La planificación semanal acordada es la siguiente:

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Paquete de tareas** | **Semana** | | | | | |
| **Nombre** | **16-04** | **23-04** | **30-04** | **07-05** | **14-05** | **21-05** |
| Lanzamiento | X |  |  |  |  |  |
| Redacción de la documentación |  | X | X | X | X | X |
| Búsqueda de conjunto de datos |  | X |  |  |  |  |
| Objetivo P-01 |  |  |  |  | X |  |
| Objetivo S-01 |  |  | X | X |  |  |
| Objetivo S-02 |  |  | X | X |  |  |
| Objetivo S-03 |  |  | X | X |  |  |

## ANÁLISIS DE LA VIABILIDAD

<Analizar el coste-beneficio. El coste abarcará tanto en tiempo, como en personal y en tecnología e infraestructuras.>

## RIESGOS

<Indicar qué pasaría si no se alcanzan los objetivos en cada uno d ellos paquetes de trabajo/tareas> <Explicar las tecnologías usadas en cada etapa de la solución de IN desarrollada> [1 párrafo]

# ANÁLISIS

## ESTABLECIMIENTO DE LOS REQUISITOS DEL SISTEMA

<Recoger todos los requisitos especificados por el usuario que deba cumplir la aplicación en el momento de abordarla, plasmándolos en un documento de requerimientos>

<Definición, análisis y validación de los requisitos. El objetivo de esta actividad es obtener un catálogo detallado de los requisitos, a partir del cual se pueda comprobar que los productos generados se ajustan a los requisitos de usuario.>

Los requisitos funcionales de la solución son los siguientes:

|  |  |
| --- | --- |
| **RF-01** | **Carga de datos automatizada desde la API de NYPD** |
| **Autores** | Juan Antonio Rodríguez, Alberto Cárdenas, Manuel Ridao |
| **Fuentes** | Necesidad del usuario |
| **Objetivos asociados** | * **OBJ P-01**: Optimizar la ubicación de las comisarías de policía * **OBJ S-01**: Visualizar las zonas más conflictivas y más seguras de la ciudad * **OBJ S-02**: Visualizar los datos en tiempo real con un cuadro de mandos * **OBJ S-03**: Diseñar un modelo predictivo para detectar la ofensa cometida |
| **Descripción** | El sistema deberá contemplar las siguientes funcionalidades:   * Conexión desde Pentaho a la API de NYPD * Petición de datos a esta API mediante REST * Transformación de los datos para su uso posterior * Carga de datos procesados en varios ficheros |

|  |  |
| --- | --- |
| **RF-02** | **Cálculo y muestra de ubicaciones óptimas de comisarías** |
| **Autores** | Juan Antonio Rodríguez, Alberto Cárdenas, Manuel Ridao |
| **Fuentes** | Necesidad del usuario |
| **Objetivos asociados** | * **OBJ P-01**: Optimizar la ubicación de las comisarías de policía |
| **Descripción** | El sistema deberá contemplar las siguientes funcionalidades:   * Clustering de ofensas según su ubicación geográfica. * Cálculo de los centroides de los clusters * Visualización de los resultados en un mapa * Mostrar el mapa a través del navegador |

|  |  |
| --- | --- |
| **RF-03** | **Visualización de un mapa de calor de delincuencia** |
| **Autores** | Juan Antonio Rodríguez, Alberto Cárdenas, Manuel Ridao |
| **Fuentes** | Necesidad del usuario |
| **Objetivos asociados** | * **OBJ S-01**: Visualizar las zonas más conflictivas y más seguras de la ciudad |
| **Descripción** | El sistema deberá contemplar las siguientes funcionalidades:   * Visualización a través de un mapa de calor de las zonas más conflictivas de la ciudad * Mostrar dicho mapa a través del navegador |

|  |  |
| --- | --- |
| **RF-04** | **Modelo predictivo para el tipo de ofensa cometida** |
| **Autores** | Juan Antonio Rodríguez, Alberto Cárdenas, Manuel Ridao |
| **Fuentes** | Necesidad del usuario |
| **Objetivos asociados** | * **OBJ S-03**: Diseñar un modelo predictivo para detectar la ofensa cometida |
| **Descripción** | El sistema deberá contemplar las siguientes funcionalidades:   * Construir un modelo predictivo para que se pueda predecir el tipo de ofensa cometida a través de los datos de entrada * Visualizar la fidelidad de dicho modelo a través de un navegador |

## ANÁLISIS DE CASOS DE USO

<Establecer los distintos casos de uso de la aplicación>

<Especificar cómo se resuelven estos a través de la interacción de los distintos actores.>

Los casos de uso de la aplicación se recogen en el siguiente diagrama



El material necesario para resolver los casos de uso es el siguiente:

|  |  |  |
| --- | --- | --- |
| **Identificador** | **Caso de uso** | **Material** |
| CU01 | ETL de los datos | Pentaho Data Integration |
| CU02 | Conexión con API | Pentaho Data Integration |
| CU03 | Visualizar mapa de calor | Servidor Web y navegador |
| CU04 | Cálculo de ubicación de comisarías de policía | KNIME Analytics Platform |
| CU05 | Visualizar la ubicación de comisarías de policía | Servidor Web y navegador |
| CU06 | Generar modelo predictivo | KNIME Analytics Platform |
| CU07 | Visualizar resultados de modelo predictivo | Servidor Web y navegador |

## ESPECIFICACIÓN DEL PLAN DE PRUEBAS

Para las pruebas de integración e implantación, se probarán en sucesión las funcionalidades de la aplicación.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Pruebas de Integración e Implantación** | | | | | |
| **Id** | **Pasos que seguir** | **Datos de entrada** | **Salida esperada** | **OK?** | **Observaciones** |
| 1 | Lanzar la ETL con Pentaho | Ninguno | Los ficheros CSV:   * list\_of\_offense\_types * report\_map\_dataset * offense\_groupby\_boro * offense\_groupby\_boro\_type * complete\_dataset | OK | Se cargan los datos de la API y se guardan en ficheros. El proceso es lento porque hay muchos datos |
| 2 | Lanzar el servidor web y acceder a la URL de la aplicación | Ninguno | Ninguno | OK |  |
| 3 | Lanzar la construcción del modelo predictivo | Los ficheros CSV:   * complete\_dataset | Los ficheros CSV:   * confusión\_matrix * cluster\_centroids | OK | El proceso es lento porque hay muchos datos |
| 4 | Lanzar el clustering de ofensas | Los ficheros CSV:   * complete\_dataset * nypd\_precincts | Los ficheros CSV:   * clustered\_results * cluster\_centroids | OK | El proceso es lento porque hay muchos datos |
| 5 | Visualizar los datos en el navegador | Todos los ficheros CSV | Ninguna | OK | El proceso es lento porque hay muchos datos |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Pruebas del sistema** | | | | |
| **ID** | **Descripción** | **Resultado** | **OK?** | **Observaciones** |
| 1 | Probar el sistema en distintos equipos | El sistema se comporta del mismo modo en distintos equipos | OK | - |
| 2 | Probar la parametrización del sistema | La parametrización funciona correctamente | OK | KNIME Analytics no permite la parametrización de rutas, por lo que es necesaria su introducción de manera manual |

# DISEÑO

## CAPACIDAD DE MEMORIA DE LA ORGANIZACIÓN

## CAPACIDAD DE INTEGRACIÓN DE INFORMACIÓN

## CAPACIDAD DE CREAR CONOCIMIENTO

## CAPACIDAD DE PRESENTACIÓN

# IMPLEMENTACIÓN

# DESPLIEGUE

# CONCLUSIONES

1. <https://www1.nyc.gov/site/nypd/about/about-nypd/about-nypd-landing.page> [↑](#footnote-ref-1)
2. <https://mypolicyviews.wordpress.com/2014/01/09/stop-and-frisk-and-the-new-york-city-murder-rate/> [↑](#footnote-ref-2)
3. <https://opendata.cityofnewyork.us/> [↑](#footnote-ref-3)