

## Caso de Estudio NYC Fire



En el siguiente caso de estudio aplique los conceptos estudiados (tomado de Big Data, la revolución de los datos masivos, de Viktor Mayer-Schönberger y Kenneth Cukier):

“...Unos cuantos años después, ya de vuelta en Nueva York, Flowers comprendió que esos métodos indicaban una forma más potente de combatir el crimen que todas las que había tenido a su disposición como Fiscal. Y halló una auténtica alma gemela en el Alcalde la Ciudad, Michael Bloomberg, quien había hecho fortuna con los datos, suministrando información financiera a la Banca. En 2009, Flowers fue asignado a una unidad especial encargada de procesar datos que ayudaran a desenmascarar a los villanos del escándalo de las Hipotecas *subprime*. La unidad tuvo tanto éxito que, un año más tarde, el Alcalde Bloomberg, le pidió que ampliara su campo de actuación.

Flowers se convirtió en el primer Director Analítico de la ciudad. Su misión: construir un equipo con los mejores científicos de datos que pudieran encontrar y explotar los montones de información virgen de la ciudad para aumentar la eficiencia en todos los terrenos posibles.

Flowers buscó a fondo hasta dar con las personas adecuadas. *“No tenía ningún interés en estadísticos con mucha experiencia –explica–. Me preocupaba bastante que se mostraran reticentes a adoptar este enfoque novedoso de la resolución de problemas”*. Anteriormente, cuando había entrevistado a varios estadísticos para el proyecto sobre el fraude financiero, estos habían mostrado cierta tendencia a expresar preocupaciones abstrusas<sup>1</sup> acerca de los modelos matemáticos. *“Yo ni siquiera pensaba en qué modelo iba a usa. Quería percepciones con las que poder actuar, eso era lo único que me importaba”*, dice. Al final, escogió un equipo de cinco personas a los que llama “Los chicos”. Todos menos uno eran recién licenciados en economía, salidos de la universidad hacía solo uno o dos años, sin mucha experiencia sobre la vida en una gran ciudad, y todos tenían un acusado lado creativo.

---

<sup>1</sup> De difícil comprensión (Nota del Profesor).

Entre los primeros desafíos a los que se enfrentó las “conversiones ilegales”: la práctica de subdividir un alojamiento en muchas unidades más pequeñas para acabar acomodando hasta 10 veces más personas que lo proyectado. Esta clase de viviendas supone un gran riesgo de incendios, además de ser foco de delitos, drogas, enfermedades y de plaga de insectos. Por las paredes puede que culebree un lío de alargadores de cable; suele haber infernillos<sup>2</sup> eléctricos en equilibrios inestables sobre los cabeceros de las camas. La gente que vive hacinada de este modo corre un gran riesgo de perecer en incendios. En el año 2005, dos bomberos murieron al intentar rescatar a unos ciudadanos. La Ciudad de Nueva York recibe aproximadamente unas 25.000 quejas anuales por conversiones ilegales, pero solo dispone de 200 Inspectores para investigarlas. No parecía haber ningún sistema para distinguir los casos meramente molestos, de los que estaban a punto de estallar. Para Flowers y sus chicos, sin embargo, este problema iba a poder resolverse con muchos datos. Empezaron con una lista de todas las propiedades de la ciudad: las 900.000 que hay. A continuación, le agregaron conjuntos de datos procedentes de 19 organismos distintos que indicaban, por ejemplo, si el propietario del inmueble no pagaba los impuestos inmobiliarios, si había habido ejecuciones hipotecarias<sup>3</sup>, y si alguna anomalía en el consumo o falta de pago de los servicios había supuesto algún tipo de corte de los mismos. También incorporaron información sobre la clase de edificios y su fecha de construcción, amén de visitas de ambulancia, tasas de delitos, quejas por roedores y demás. Luego, compararon toda esta información con 5 años de datos sobre incendios clasificados por gravedad y buscaron correlaciones<sup>4</sup> para intentar generar un sistema que permitiera predecir que quejas deberían ser atendidas con la mayor urgencia.

Al principio, buena parte de los datos no estaban en un formato aprovechable. Por ejemplo, los responsables de los archivos de la ciudad no tenían una forma única y normalizada de describir la localización; cada organismo y cada departamento parecían usar su propio sistema. El Departamento de Inmuebles le asigna un número único a cada estructura, pero el de Mantenimiento de la Vivienda emplea un sistema de numeración diferente. Por su parte el Departamento de Hacienda le atribuye a cada propiedad una referencia basada en el distrito municipal, la manzana y la parcela mientras que la Policía utiliza coordenadas cartesianas<sup>5</sup>. Los Bomberos se basan en un sistema de proximidad a las cabinas de teléfonos de emergencia en relación con el emplazamiento de sus cuarteles, aun cuando esas cabinas ya no existan. Los chicos de Flowers abarcaron este desorden inventando un sistema de identificación de edificios: usaron un área pequeña de la fachada de cada casa basada en coordenadas cartesianas y luego importaron otros datos de geolocalización de las bases de datos de las demás organizaciones. Su método era inherentemente inexacto pero la ingente<sup>6</sup> cantidad de datos que eran capaces de usar compensaban con creces las imperfecciones.

El equipo, sin embargo, no se contentó con procesar datos. Salieron al terreno con los Inspectores para ver como trabajaban. Tomaron copiosas<sup>7</sup> notas y les hicieron todo tipo de preguntas a los profesionales. Cuando un canoso Jefe de Inspectores decía refunfuñando que el edificio que estaba a

<sup>2</sup> Se refiere a hervidores y calentadores (Nota del Profesor).

<sup>3</sup> Remate de propiedades por no pago de su deuda (Nota del Profesor).

<sup>4</sup> Correspondencia o relación recíproca entre dos o más cosas o series de cosas (Diccionario RAE) (Nota del Profesor).

<sup>5</sup> Sistema de referencia que normalmente divide un terreno en dos dimensiones registrando la ubicación de una propiedad en la intersección de dos puntos (Nota del Profesor).

<sup>6</sup> Cantidad muy grande (Nota del Profesor).

<sup>7</sup> Numerosa, abundante (Nota del Profesor).

Docente Diseñador	Gerardo Cerda Neumann	Revisor metodológico	Manuela Jiménez
-------------------	-----------------------	----------------------	-----------------

punto de visitar no sería un problema los chicos quisieron saber por qué estaba tan seguro. El hombre no supo explicárselo del todo, pero los chicos determinaron gradualmente que su intuición se basaba en que el edificio tenía la fachada renovada hacía poco, lo que sugería que el propietario se preocupaba por el sitio.

Los chicos regresaron a sus cubículos y se preguntaron cómo podrían introducir “renovado reciente” en su modelo, a guisa de señal. Al fin y al cabo, los ladrillos no están *datificados*... aún. Pero, por supuesto, es preceptiva<sup>8</sup> una autorización municipal para realizar cualquier trabajo en la fachada.

Añadir la información sobre los permisos mejoró las prestaciones predictivas del sistema al indicar que algunas propiedades bajo sospecha probablemente no constituyeran riesgos mayores.

La analítica también mostraba a veces que ciertas maneras que hacer las cosas, consagradas por el tiempo, no eran las mejores, igual que aquellos ojeadores de *Moneyball*<sup>9</sup> habían tenido que aceptar las deficiencias de su intuición. Por ejemplo, el número de llamadas al 311, el teléfono de quejas urgentes de la ciudad se consideraba indicativo de qué edificios estaban más necesitados de atención: más llamadas equivalían a un problema más grave. Pero esta resultó ser una medida que inducía a error. Una sola rata detectada en el Upper East Side<sup>10</sup>, el barrio *Chic*, podía generar 30 llamadas en el espacio de una hora, pero hacía falta un batallón entero de roedores antes que los residentes del Bronx<sup>11</sup> se sintieran impelidos a marcar el 311. De igual modo, la mayoría de las quejas por una conversión ilegal podían estar relacionadas con el ruido, no con las situaciones de riesgo.

En junio de 2011, Flowers y sus chicos pusieron en marcha su sistema. Todas las quejas que caían en la categoría de conversión ilegal se procesaban semanalmente. Reunieron todas las que quedaron clasificadas en el 5% superior, como las de mayor riesgo de incendio y se las pasaron a los Inspectores para su inmediata investigación. Cuando llegaron los resultados, todo el mundo se quedó asombrado.

Antes del análisis de datos masivos los inspectores investigaban las quejas que les parecían de peor agüero, pero solo en el 13% de los casos hallaban condiciones lo bastante graves para requerir una orden de desalojo del inmueble. Ahora cursaban órdenes desalojo en más del 70% de los edificios que inspeccionaban. Determinando así que edificios requerían más urgente su atención, los datos masivos multiplicaron por 5 la eficiencia de la inspección. Y su trabajo se convirtió además más satisfactorio: se concentraban en los problemas más graves. La eficacia redoblada de los Inspectores tuvo asimismo beneficios indirectos. Los incendios en las conversiones ilegales suponen una posibilidad 15 veces mayor de provocar heridos o muertos entre los Bomberos que intervienen, por lo que el Departamento de Bomberos se mostró encantado. Flowers y sus chicos parecían magos con una bola de cristal que les permitía ver el futuro, y predecir que sitios presentaban el mayor riesgo. Tomaron cantidades masivas de datos que llevaban años tirados, en su mayor parte sin usar desde su recogida, y los explotaron de forma novedosa para sacarle valor real. Usar un gran *corpus* de información les

<sup>8</sup> Obligatoria por una norma administrativa (Nota del Profesor).

<sup>9</sup> Famoso libro de análisis de datos del béisbol que le permitió a un equipo aumentar la precisión al contratar nuevos jugadores analizando las estadísticas de juego de los postulantes (Nota del Profesor).

<sup>10</sup> Sector lujoso de la ciudad que está junto al Central Park (Nota del Profesor).

<sup>11</sup> Sector norte de la ciudad donde existen muchas propiedades de bajo y costo (Nota del Profesor).

Docente Diseñador	Gerardo Cerda Neumann	Revisor metodológico	Manuela Jiménez
-------------------	-----------------------	----------------------	-----------------

permitió advertir conexiones que no eran detectables en cantidades más pequeñas: es la esencia de los datos masivos.

... Las sospechas de los expertos tuvieron que ceder protagonismo ante el enfoque basado en datos. Al mismo tiempo, Flowers y sus chicos sometieron continuamente a prueba su sistema con los Inspectores veteranos, aprovechando su experiencia para mejorar el funcionamiento. Pero, con todo, la razón principalmente del éxito del programa fue que prescindió de la causalidad en favor de la correlación.

*“No me interesa la causalidad salvo en la medida en que lleva a la acción – explica Flowers –. La causalidad es para otras personas y francamente se me antoja muy arriesgado empezar a hablar de causalidad. No creo que exista ni una sola causalidad entre el día en que alguien presenta una demanda de ejecución hipotecaria sobre una propiedad y el que esta finca tenga o no un riesgo histórico de incendio estructural. Me parece que sería obtuso pensarlo. Y, de hecho, nadie saldría a decir eso. Pensarían: no, son los factores subyacentes. Pero ni siquiera quiero entrar en eso. Necesito un punto de datos específicos al que pueda acceder, y saber su importancia. Si es significativo actuaremos en consecuencia. Si no lo es, no haremos nada. Mire, tenemos auténticos problemas que resolver. Sinceramente, no puedo andar perdiendo el tiempo pensando en otras cosas como la causalidad, ahora mismo”.*

---