

Meta: DATA_

Plataforma de competencias en Ciencia de Datos

[Volver a la página de inicio](#)

Predicción de ventas estacionales

La competencia finaliza el 18 de Julio de 2019, quedan aún 23 días para participar.

Para participar, [Subí tu respuesta](#) .

Introducción

En AlixPartners trabajamos con diferentes problemas de negocio que requieren un análisis de los datos para extraer conclusiones. A continuación se describe el problema a resolver

"Fiestas Dulces" es una empresa que entrega sus artículos en los puntos de venta a terceros (consignación). El producto insignia de la empresa se consume mayoritariamente en las fiestas a fin de año. El problema es que los artículos que comercializa son perecederos. La práctica usual es descartar lo que no se vende.

En la actualidad la empresa goza de buenos ingresos, pero considerando el descarte de artículos el margen podría mejorar. Para maximizar la ganancia "Fiestas Dulces" necesita minimizar dicho descarte. Una forma de hacer esto es conseguir buenas predicciones sobre la venta de sus productos en cada punto de venta.

El objetivo de esta competencia es predecir la cantidad de artículos que se venderá en cada punto de venta en diciembre de 2014.

Premios

Primer puesto:

- 2x auriculares inalámbricos [Bose QuietComfort](#).
- 2x altavoz [Echo Plus](#)

Segundo Puesto:

- 2x altavoz [Echo Plus](#)

Esta competencia se organiza en el marco [33ra Escuela Ciencias Informáticas \(ECI\)](#) que se llevará a cabo en Buenos Aires , del 22 al 26 de Julio de 2019. La competencia termina en la semana anterior, y la entrega de premios será durante la ceremonia de cierre de la Escuela. Los premios son gentileza de **AlixPartners** y serán entregados a los ganadores de manera directa por la Empresa.

Tabla de posiciones

Puntajes hasta ahora.

Usuario

damianealeman

sergio.chouhy

martinehman90

federicovms

Evaluación

La calificación de la competencia se logra en un 85% con el resultado de las predicciones y un 15% con un informe en donde los participantes expliquen qué tratamiento hicieron de los datos, y qué decisiones tomaron para el modelado. Se leerán todos los informes presentados, incluso aquellos con bajo puntaje en las predicciones. **¡A no desalentarse!**

La calificación de la solución propuesta se hace con el error absoluto medio ([MAE](#) por su sigla en inglés) y se calcula como el promedio de las diferencias (en valor absoluto) entre las respuestas enviadas y las correctas:

$$\text{MAE} = \frac{1}{n} \sum_{j=1}^n |y_j - \hat{y}_j|$$

Como en todas las competencias, cada solución propuesta es evaluada en contraposición a las respuestas correctas. El conjunto de respuestas se divide en dos, de manera secreta y al azar. Durante la competencia se mostrará el puntaje correspondiente a una mitad de puntos de venta posibles. Al finalizar la competencia, el puntaje válido será el correspondiente a la *otra* mitad de puntos de venta. **¡A no sobreajustar!**

En todo momento el puntaje válido corresponderá al **último** intento realizado.

El 15% que resta de la nota lo otorga un jurado ad-hoc de especialistas en el tema evaluando el informe sobre el trabajo realizado con el dataset. El jurado podrá requerir reproducir el puntaje correspondiente al último intento registrado en la competencia.

Las predicciones se deben enviar en un archivo CSV sin encabezado conteniendo solamente la identificación del POS (id_pos) y la columna de predicciones (unidades). Las predicciones pueden contener números enteros solamente, las filas que no respeten esto no se considerarán. A continuación, se muestran cinco filas de un archivo de predicciones típico:

```
3142, 12
3143, 24
3144, 32
3145, 45
3146, 54
```

Los id_pos (primer columna) deben coincidir con los id_pos del archivo [Ejemplo de Respuesta](#), en caso contrario la plataforma devolverá un error

Bueno, ya ¡dame la data!

- [Envíos](#)
- [Puntos de Venta](#)
- [Ventas](#)
- [Ejemplo de Respuesta](#)

Descripción

Los participantes disponen de tres archivos que reúnen la información del caso, envios.txt, ventas.txt y pos.csv.

envios.txt:

- *fecha*: Fecha de envío.
- *id_pos*: Identificador único de Punto de Venta.
- *unidades_despachadas*: Cantidad de productos enviados al POS en la fecha mencionada.
- *cantidad_envios*: Cantidad de envíos realizados para la fecha y POS en cuestión.

ventas.txt:

- *fecha*: Fecha de venta.
- *id_pos*: Identificador único de Punto de Venta.
- *cantidad_envios*: Cantidad de productos vendidos.
- *canal*: Es la clase de tienda en la que se realizó la venta.

pos.csv:

- *id_pos*: Identificador único de Punto de Venta.
- *competidores*: Número de competidores en un radio de una milla del POS.
- *ingreso_mediana*: Mediana del ingreso para el código postal en cuestión.
- *ingreso_promedio*: Ingreso promedio para el código postal en cuestión.
- *densidad_poblacional*: Densidad poblacional para el código postal en cuestión.
- *pct_0a5*: Porcentaje de la población de 0 a 5 años.
- *pct_5a9*: Porcentaje de la población de 5 a 9 años.
- *pct_10a14*: Porcentaje de la población de 10 a 14 años.
- *pct_15a19*: Porcentaje de la población de 15 a 19 años.
- *pct_20a24*: Porcentaje de la población de 20 a 24 años.
- *pct_25a29*: Porcentaje de la población de 25 a 29 años.
- *pct_30a34*: Porcentaje de la población de 30 a 34 años.
- *pct_35a39*: Porcentaje de la población de 35 a 39 años.
- *pct_40a44*: Porcentaje de la población de 40 a 44 años.
- *pct_45a49*: Porcentaje de la población de 45 a 49 años.
- *pct_50a54*: Porcentaje de la población de 50 a 54 años.
- *pct_55a59*: Porcentaje de la población de 55 a 59 años.
- *pct_60a64*: Porcentaje de la población de 60 a 64 años.
- *pct_65a69*: Porcentaje de la población de 65 a 69 años.
- *pct_70a74*: Porcentaje de la población de 70 a 74 años.
- *pct_75a79*: Porcentaje de la población de 75 a 79 años.
- *pct_80a84*: Porcentaje de la población de 80 a 84 años.
- *pct_85ainf*: Porcentaje de la población mayor a 85 años.
- *pct_bachelors*: Porcentaje de la población con título de bachelor.
- *pct_doctorados*: Porcentaje de la población con título de doctor.
- *pct_secundario*: Porcentaje de la población con título secundario.
- *pct_master*: Porcentaje de la población con título de master.
- *pct_bicicleta*: Porcentaje de utilización de bicicletas.
- *pct_omnibus*: Porcentaje de utilización de omnibus.
- *pct_subtes*: Porcentaje de utilización de subtes.
- *pct_taxi*: Porcentaje de utilización de taxis.
- *pct_caminata*: Porcentaje de transporte de a pie.
- *mediana_valor_hogar*: Mediana de valor del hogar en la vecindad del POS.