

PRESENTACION EDA SOBRE INCIDENCIAS EN LINEA DE PRODUCCION

Objetivo principal

Los objetivos principales son:

- Localizar los focos donde el coste de reparación es mayor.
- Buscar un patrón donde las incidencias son mayores

Datos

Los datos están muy bien estructurados y no contienen datos nulos.

Solo hemos convertido los datos tipos fecha de object a datetime.

Esta es la clasificación por tipo de variable

Columna/Variable	Descripción	Tipo_de_Variable
defect_id	Identificador único de cada defecto	Indice único
product_id	Identificador del producto	Numerica Discreta
defect_type	Tipo de defecto	Categorica
defect_date	Fecha en la que se detectó el defecto	Fecha
defect_location	Ubicación donde se detectó el defecto	Categorica
severity	Nivel de gravedad	Categorica
inspection_method	Método utilizado para detectar el defecto	Categorica
repair_cost	Coste de reparación del defecto	Numerica continua

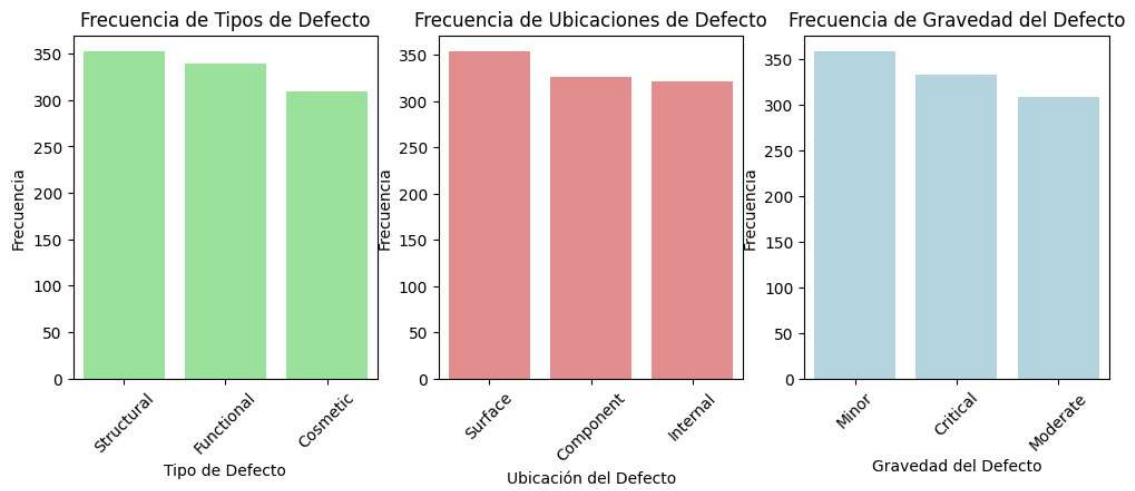
Se aprecia que los datos puede ser categorizados por 3 variables (Gravedad, Ubicación y Tipo) estas tres categorías, a su vez tiene tres valores únicos, La siguiente tabla muestra las opciones que tenemos y vamos a realizar los análisis pertinentes para cumplir con los objetivos

Tipo	Ubicación	Gravedad
Structural (Estructural)	Surface (Superficie)	Minor (Menor)
Functional (Funcional)	Component (Componente)	Critical (Critico)
Cosmetic (Estético)	Internal (interno)	Moderate (Moderado)

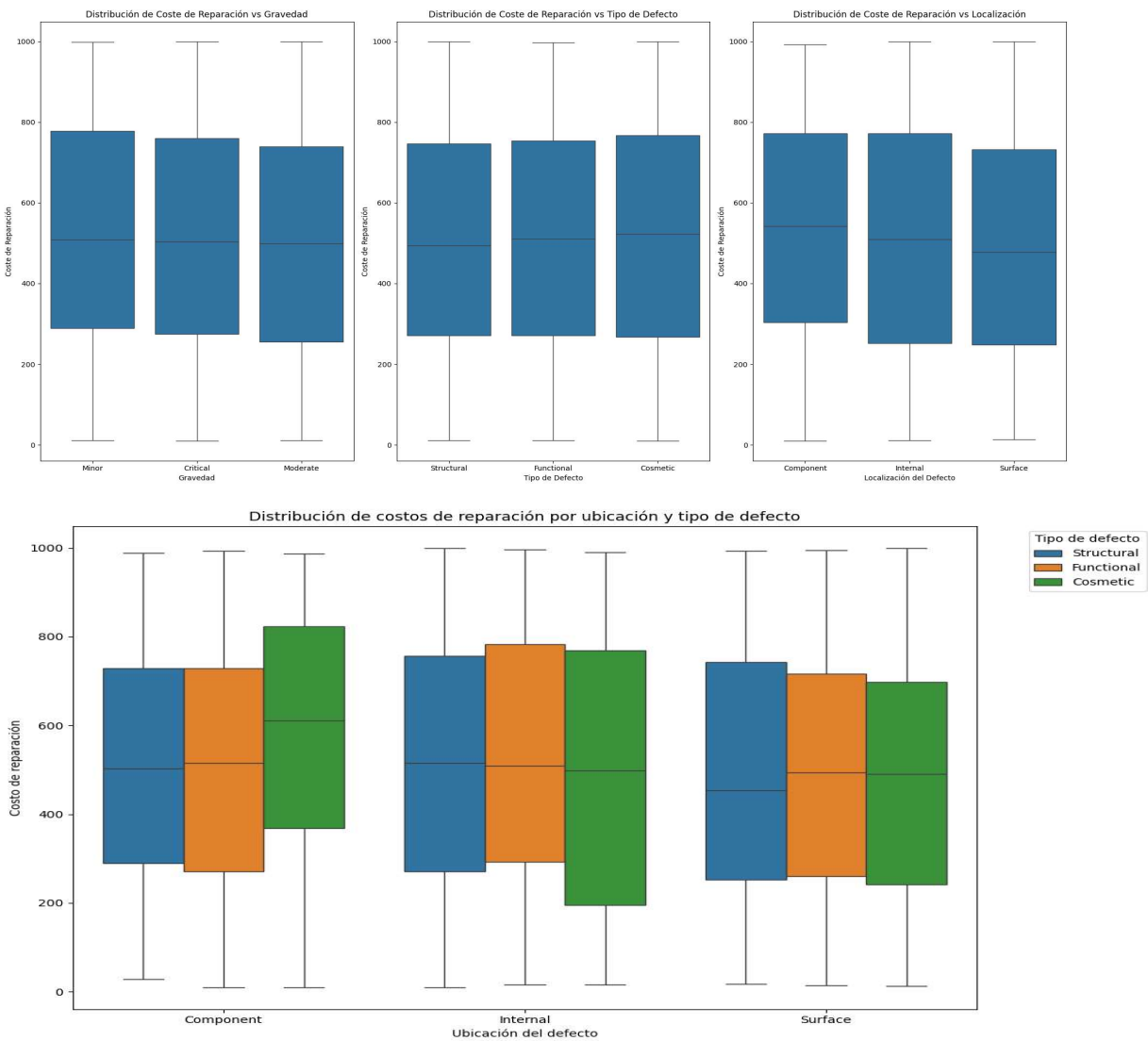
Proceso de Análisis

Primero vamos a buscar la frecuencia para cada una de las variables (Tipo de defecto, ubicación y Gravedad) para ver si hemos de prestar más atención a un alguna de las variables categóricas. Las

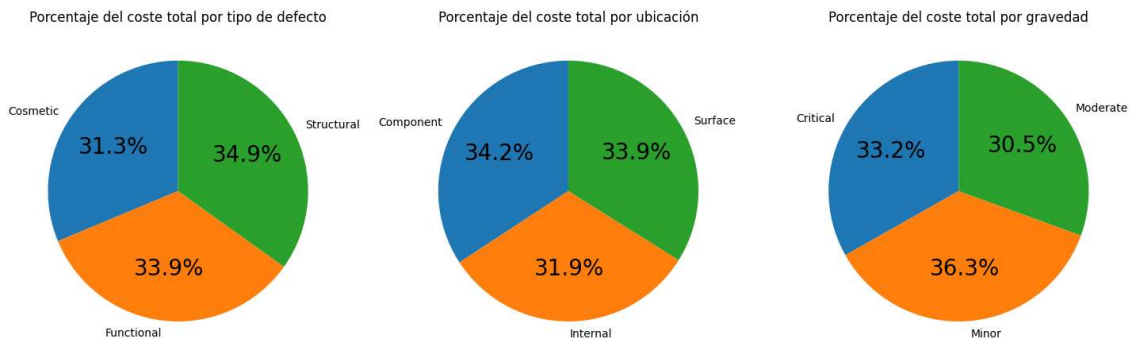
siguientes gráficas muestra unos datos bastante homogéneos, no hay grandes diferencias entre los datos.



El siguiente paso que he realizado es ver la distribución del coste de reparación para las variables categóricas, el resulta esta en la misma línea que los resultados anteriores. Distribución equitativa sin outliers.



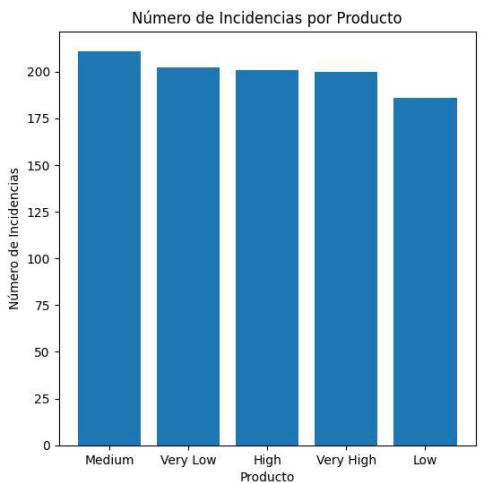
He representado en gráficos circulares el porcentaje de del coste para variable de total del dataframe. Para ello he creado una función que después he utilizado para profundizar más según el coste de reparación, lo veremos en el siguiente paso.



Como hemos visto los valores vuelven a estar distribuidos de forma equitativa. vamos a clasificar los datos por el coste de reparación. De más a menos importe

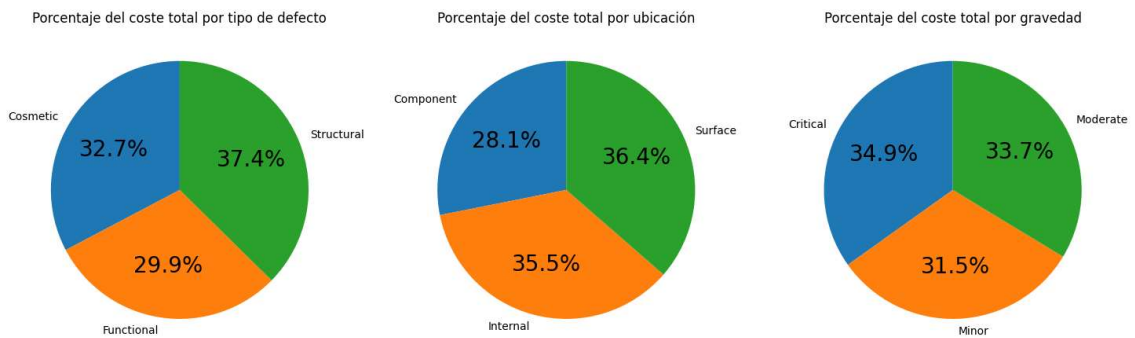
Coste Muy Bajo, Bajo, Medio, Alto, Muy alto.

El número de incidencias vuelve a estar distribuido de forma equitativa.

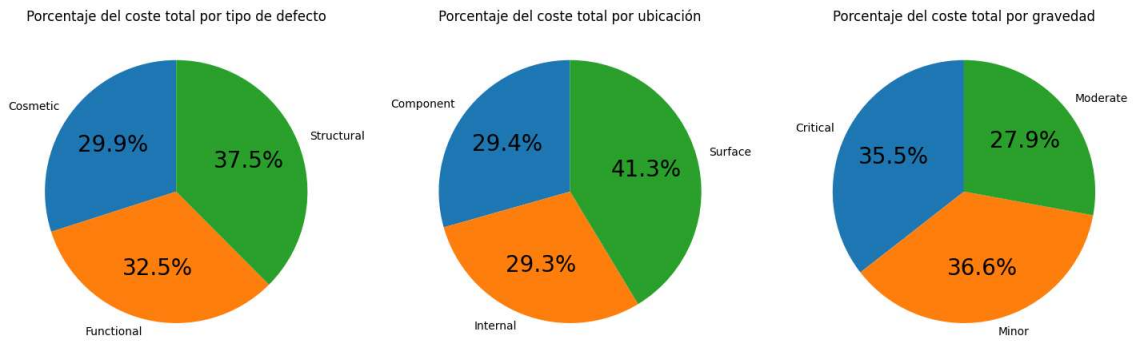


He mostrado mediante los gráficos circulares para el porcentaje para cada una de las categorías y tipo de coste, no ha sido un trabajo excesivo por que he aprovechado la función creada para el punto anterior.

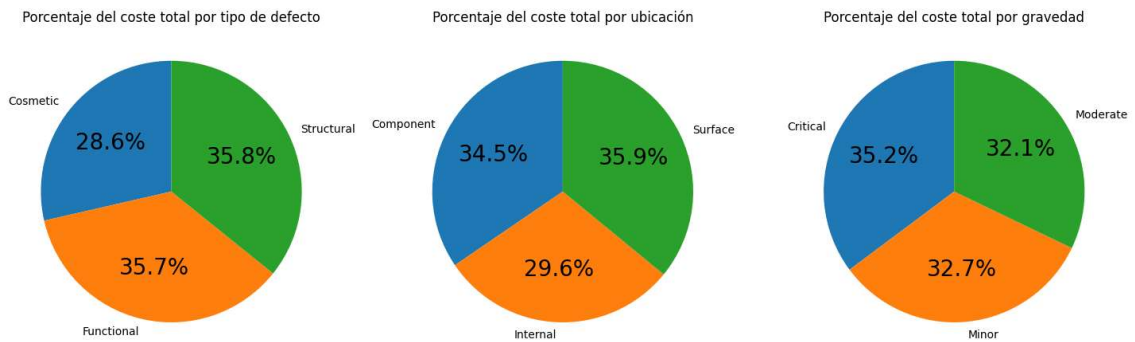
Muy Bajo



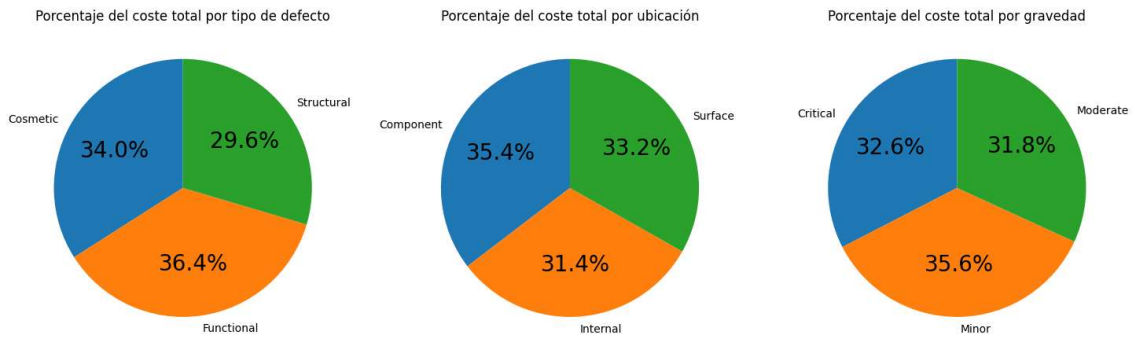
Bajo



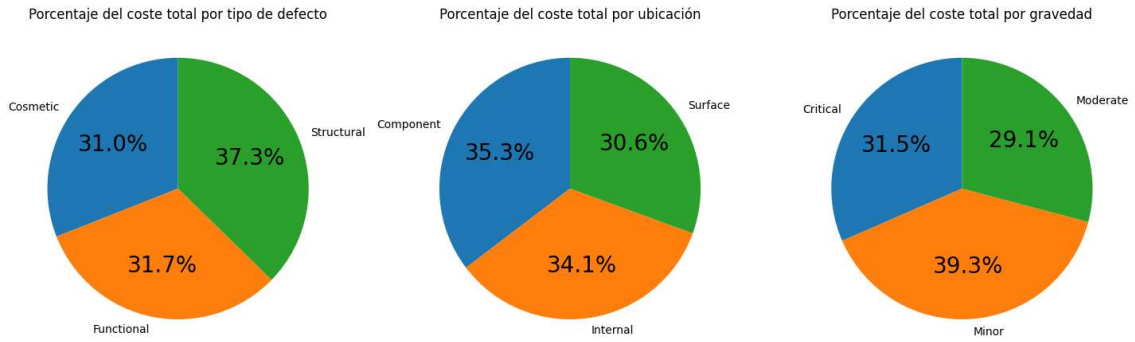
Medio



Alto



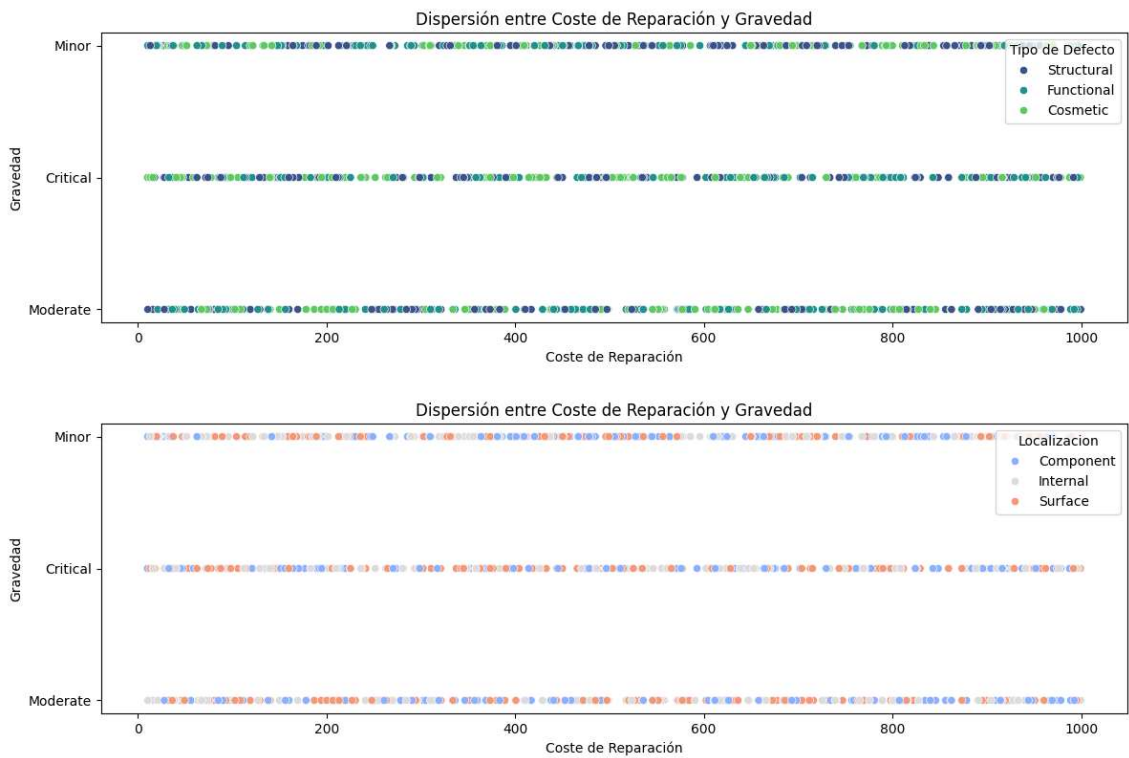
Muy Alto



El siguiente paso ha sido analizar la dispersión de los costes.

El resultado ha sido una dispersión muy alta y a su vez muy repartida.

Los valores van de 10 a 1000 aproximadamente, pero en todas las categorías están repartidos de forma muy similar, a lo largo de todo el rango de importes, sucede esto para las tres variables categóricas (Gravedad, tipo de defecto, Localización)

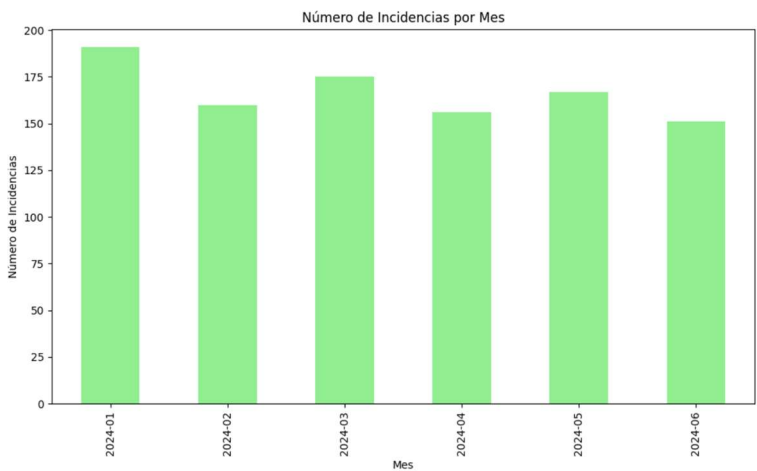


Mediante el calculo de χ^2 he buscado correlaciones entre las variables categóricas

No hay correlación entre estas variables Gravedad – Tipo defecto – Localización.

Dado los resultados, he hecho un último análisis sobre el tiempo, analizando el número defectos que ha habido por cada mes.

Los resultados no presentan sorpresas, continuamos con un reparto equitativo de los datos:



Conclusión

Tanto los tipos de defecto, la gravedad y la localización se distribuyen de manera similar en cuanto a frecuencia y coste de reparación. No hay desviaciones importantes a tener en cuenta. Esto puede reflejar que la producción se realiza de forma homogénea, pero también podría reflejar una falta de variables a la hora de clasificar las incidencias. (p. ejemplo más detalle en el tipo de incidencia)

La dispersión del coste de reparación es muy alta, pero está muy repartida en todos los aspectos.

Como medida para profundizar más en el análisis, se puede detallar más las variables aumentando la cantidad de las mismas si el campo lo permite.

Sería bueno realizar este análisis de forma periódica por ejemplo de forma mensual, trimestral, semestral, anual, para confirmar estos datos.