

Informe de Análisis de Datos - Ventas de Limonada

Introducción

Este informe presenta un análisis detallado del dataset "Lemonade.txt", que contiene datos sobre las ventas diarias de limonada en 2017. Se analizaron factores como la temperatura, la lluvia, la distribución de volantes publicitarios y su relación con las ventas diarias. Los resultados obtenidos permitirán entender mejor las dinámicas que influyen en las ventas y tomar decisiones informadas para optimizar la estrategia de negocio.

Descripción del Dataset

El dataset "Lemonade.txt" contiene 365 registros, uno por cada día del año 2017. Las columnas del dataset son las siguientes:

- **Date:** Fecha del día en formato DD/MM/AAAA.
- **Día:** Día de la semana (Sunday, Monday, etc.).
- **Temperature:** Temperatura promedio del día.
- **Rainfall:** Precipitación diaria.
- **Flyers:** Volantes distribuidos ese día.
- **Price:** Precio unitario de la limonada (dólares).
- **Sales:** Limonadas vendidas ese día.

El análisis busca responder preguntas clave sobre el impacto de las condiciones climáticas y publicitarias en las ventas, así como la identificación de patrones y valores atípicos en las variables.

Análisis Estadístico

Este análisis te permitirá entender mejor la distribución de los datos, la presencia de posibles valores atípicos y cómo se relacionan las variables clave con las ventas.

1. Medidas de Tendencia Central y Dispersión

Para las variables numéricas del dataset, se calcularon las siguientes medidas:

- **Media:** Valor promedio de cada variable.
- **Mediana:** Valor central en la distribución.
- **Desviación estándar:** Medida de la dispersión en torno a la media.
- **Varianza:** Medida de la variabilidad de los datos.
- **Percentiles:** Valores de referencia en el 25%, 50% (mediana), y 75%.

2. Detección de Valores Atípicos (Outliers)

Utilizando el rango intercuartílico (IQR), se identificaron valores atípicos en algunas variables. Los outliers son puntos que caen fuera del rango de $[Q1 - 1.5 \cdot IQR, Q3 + 1.5 \cdot IQR]$.

Resultados:

- **Temperatura:** Se identificaron 8 outliers en los días con temperaturas extremadamente bajas o altas.
- **Lluvia:** Se detectaron 28 outliers correspondientes a días con precipitaciones inusuales.
- **Flyers:** Dos outliers se encontraron en días con volantes distribuidos en cantidades significativamente mayores a la media.
- **Ventas:** No se detectaron outliers en las ventas, lo que sugiere un comportamiento estable.

Interpretación de Outliers

1. Temperatura

- **Estadísticas Descriptivas:**
 - Media: 59.61
 - Mediana: 60.80
 - Desviación Estándar: 17.99
 - Varianza: 323.74
 - Mínimo: 1.10
 - Máximo: 102.90
 - Percentiles: 25 (48.7), 50 (60.8), 75 (71.3)
 - Asimetría: -0.48 (indica una ligera inclinación hacia la izquierda)
 - Curtosis: 0.80 (distribución ligeramente aplanada)
- **Outliers Encontrados: 8**

Outliers encontrados: 8							
	Date	Día	Temperature	Rainfall	Flyers	Price	Sales
64	06/03/2017	Monday	3.4	0.77	28	0.3	24
86	28/03/2017	Tuesday	5.9	0.83	48	0.3	23
168	18/06/2017	Sunday	2.6	0.59	60	0.3	32
195	15/07/2017	Saturday	2.5	0.54	56	0.5	35
260	18/09/2017	Lunes	6.8	0.71	37	0.3	26
344	11/12/2017	Monday	4.1	1.11	33	0.3	17
356	23/12/2017	Saturday	2.4	1.11	20	0.3	18
364	31/12/2017	Sunday	1.1	2.50	9	0.3	7

Los outliers en la temperatura son días con valores extremadamente bajos (por ejemplo, 1.1 °F y 2.5 °F) que son significativamente inferiores a la media y la mediana. Estos valores pueden ser el resultado de errores de medición, condiciones climáticas inusuales o eventos extraordinarios. El impacto de estos outliers en el análisis general podría distorsionar las

tendencias en la relación entre la temperatura y las ventas, por lo que es crucial evaluarlos para entender si representan anomalías legítimas o errores de registro.

2. Lluvias (Rainfall)

- **Estadísticas Descriptivas:**
 - Media: 0.83
 - Mediana: 0.74
 - Desviación Estándar: 0.27
 - Varianza: 0.075
 - Mínimo: 0.47
 - Máximo: 2.50
 - Percentiles: 25 (0.65), 50 (0.74), 75 (0.91)
 - Asimetría: 1.85 (indica una fuerte inclinación hacia la derecha)
 - Curtosis: 5.16 (indica una distribución muy concentrada)
- **Outliers Encontrados: 28**

Outliers encontrados: 28							
	Date	Día	Temperature	Rainfall	Flyers	Price	Sales
0	01/01/2017	Sunday	27.0	2.00	15	0.3	10
1	02/01/2017	Monday	28.9	1.33	15	0.3	13
2	03/01/2017	Tuesday	34.5	1.33	27	0.3	15
5	06/01/2017	Friday	25.3	1.54	23	0.3	11
6	07/01/2017	Saturday	32.9	1.54	19	0.3	13
10	11/01/2017	Wednesday	32.6	1.54	23	0.3	12
11	12/01/2017	Thursday	38.2	1.33	16	0.3	14
12	13/01/2017	Friday	37.5	1.33	19	0.3	15
15	16/01/2017	Monday	30.6	1.67	24	0.3	12
16	17/01/2017	Tuesday	32.2	1.43	26	0.3	14
19	20/01/2017	Friday	31.6	1.43	20	0.3	12
23	24/01/2017	Tuesday	28.6	1.54	20	0.3	12
27	28/01/2017	Saturday	34.9	1.33	15	0.3	13
28	29/01/2017	Sunday	35.2	1.33	27	0.3	14
337	04/12/2017	Monday	34.9	1.54	16	0.3	13
338	12/05/2017	Tuesday	22.0	1.82	11	0.3	10
342	09/12/2017	Saturday	31.2	1.43	19	0.3	14
343	10/12/2017	Sunday	31.3	1.82	15	0.3	11
345	12/12/2017	Tuesday	33.5	1.33	22	0.3	15
346	13/12/2017	Wednesday	32.2	1.43	26	0.3	14
347	14/12/2017	Thursday	31.9	1.54	24	0.3	13
350	17/12/2017	Sunday	32.2	1.33	16	0.3	14
351	18/12/2017	Monday	30.9	1.43	27	0.3	13
354	21/12/2017	Thursday	40.5	1.33	23	0.3	15
355	22/12/2017	Friday	30.9	1.54	17	0.3	13
359	26/12/2017	Tuesday	28.9	1.43	23	0.3	13
363	30/12/2017	Saturday	30.9	1.43	22	0.3	13
364	31/12/2017	Sunday	1.1	2.50	9	0.3	7

La lluvia presenta una notable asimetría positiva, lo que sugiere que hay más días con lluvias leves y pocos días con lluvias intensas. Los outliers se concentran en días con precipitaciones significativamente más altas (hasta 2.5 unidades). Esto podría reflejar eventos climáticos extremos que afectan las ventas, ya que un aumento drástico en la lluvia puede impactar negativamente en la asistencia a la tienda y, por ende, en las ventas.

3. Flyers

- **Estadísticas Descriptivas:**
 - Media: 40.28
 - Mediana: 39.00
 - Desviación Estándar: 13.18
 - Varianza: 173.68
 - Mínimo: 9
 - Máximo: 80
 - Percentiles: 25 (31.00), 50 (39.00), 75 (49.00)
 - Asimetría: 0.30 (distribución relativamente simétrica)
 - Curtosis: -0.16 (ligeramente menos concentrada)
- **Outliers Encontrados: 2**

Outliers encontrados: 2							
	Date	Día	Temperature	Rainfall	Flyers	Price	Sales
166	16/06/2017	Friday	99.3	0.47	77	0.3	41
194	14/07/2017	Friday	92.0	0.50	80	0.5	40

Los outliers en la variable "Flyers" son días con valores de 77 y 80, que se alejan considerablemente de la media. Estos días pueden estar relacionados con campañas promocionales excepcionales que no son representativas del comportamiento habitual. Su inclusión en el análisis podría llevar a una sobreestimación del efecto de los flyers en las ventas, por lo que es recomendable examinarlos más a fondo.

4. Ventas (Sales)

- **Estadísticas Descriptivas:**
 - Media: 25.32
 - Mediana: 25.00
 - Desviación Estándar: 6.89
 - Varianza: 47.52
 - Mínimo: 7
 - Máximo: 43
 - Percentiles: 25 (20.00), 50 (25.00), 75 (30.00)
 - Asimetría: 0.02 (distribución casi simétrica)
 - Curtosis: -0.26 (ligeramente menos concentrada)

- **Outliers Encontrados: 0**

No se encontraron outliers en las ventas, lo que indica que los datos son coherentes y no hay valores atípicos que afecten el análisis. Esto sugiere que las ventas siguen un patrón relativamente normal y que los valores caen dentro de un rango esperado, lo cual es positivo para la validez del análisis de regresión.

Los outliers en las variables de temperatura y lluvia pueden ofrecer información valiosa sobre condiciones climáticas inusuales que afectan las ventas, mientras que en "Flyers" pueden reflejar estrategias promocionales excepcionales. Es esencial considerar estos outliers en el contexto del análisis para interpretar adecuadamente su impacto en las ventas y en el modelo de regresión lineal. Una revisión de estos casos ayudará a mejorar la precisión del modelo y su utilidad para la toma de decisiones comerciales.

Modelo Predictivo simple

Se construyó un modelo de regresión lineal para predecir las ventas diarias de limonada en función de la temperatura, la lluvia y la cantidad de volantes distribuidos. Este modelo busca proporcionar una herramienta que permita predecir el volumen de ventas bajo distintas condiciones climáticas.

Proceso:

1. **Modelo utilizado: Regresión lineal.**

- La regresión lineal es un modelo predictivo que nos permite entender cómo varía la variable dependiente (en este caso, las **ventas diarias**) en función de una o más variables independientes (como la **temperatura, lluvia y flyers**). Es simple de interpretar y puede ofrecer una buena base para predecir ventas en este contexto.

2. **Variables:**

- Dependiente: Ventas (Sales).
- Independientes: Temperatura (Temperature), Lluvia (Rainfall) y Volantes (Flyers).

3. **Estructura del Modelo:**

- Primero, ajustamos un modelo de regresión lineal usando los datos disponibles.
- Luego, evaluamos la calidad del modelo utilizando métricas como el R-cuadrado (R^2), que nos indica qué porcentaje de la variabilidad en las ventas está siendo explicado por las variables independientes.
- Si el modelo es adecuado, podemos usarlo para predecir las ventas futuras bajo diferentes condiciones climáticas.

4. **Entrenamiento y Evaluación:** El modelo fue entrenado con el 80% de los datos y evaluado con el 20% restante. Se utilizó el coeficiente de determinación (R^2) para medir la capacidad explicativa del modelo.

5. **Resultados:** El modelo arrojó un R^2 aceptable, indicando que una parte significativa de la variabilidad en las ventas puede explicarse por las condiciones climáticas y el número de volantes distribuidos. Este enfoque es útil porque proporciona una estimación simple pero efectiva de cómo las ventas de limonada podrían comportarse en función de diferentes variables. Además, la simplicidad del modelo permite interpretaciones claras que los usuarios del negocio pueden utilizar para mejorar su planificación y gestión de recursos.

Análisis Temporal

El análisis temporal es crucial para entender cómo las ventas y otras variables relacionadas cambian a lo largo del tiempo. Esto puede ayudar a identificar patrones estacionales y tendencias, lo que permite a las empresas ajustar sus estrategias y mejorar la toma de decisiones.

Tendencias Estacionales

- **Días de la Semana con Más Ventas:**
 - Al analizar las ventas promedio por día de la semana, se puede identificar cuáles son los días más rentables. Esto es útil para programar promociones o aumentar el inventario en días de alta demanda.
- **Relación entre la temperatura y las ventas a lo Largo del Tiempo:**
 - Un gráfico que muestra las ventas a lo largo del tiempo es fundamental para visualizar los ciclos de negocio y puede ayudar a identificar tendencias a largo plazo, así como la estacionalidad. En este caso se mostro como existe una relación en su comportamiento con respecto a la temperatura.
- **Cambios de Temperatura a lo Largo del Año:**
 - Es relevante observar cómo la temperatura media cambia durante el año y cómo esto afecta las ventas. Graficar ambas variables puede revelar patrones o correlaciones interesantes.

Este análisis temporal es útil para la planificación estratégica del negocio, permitiendo identificar periodos de alta demanda y ajustar las operaciones de marketing.

Conclusión del Informe

Este informe proporciona una visión integral del comportamiento de las ventas de limonada en relación con factores climáticos y de marketing. El análisis estadístico y el modelo predictivo permiten tomar decisiones más informadas sobre la gestión del negocio. La identificación de valores atípicos en temperatura y lluvia ofrece información adicional que puede ser útil para entender mejor las fluctuaciones en las ventas. Finalmente, el análisis temporal resalta la importancia de las variaciones a lo largo del año y cómo se pueden aprovechar para optimizar las estrategias comerciales.