

Protocolo de Recolección y Análisis de Datos

1. Objetivo

Definir un proceso estructurado para la recolección, depuración, análisis y validación de datos operacionales del servicio en la heladería Crepes & Waffles en el centro comercial Centro Chía. Estos datos serán usados para construir un modelo de simulación de eventos discretos que represente de manera válida el sistema.

2. Metodología de Recolección de Datos

2.1. Variables Capturadas

Las variables registradas se dividen en tres grandes categorías:

1. Proceso de atención y pago:

- Hora de arribo del cliente.
- Hora de inicio de atención del pedido.
- Hora de inicio de atención del pago.
- Hora de finalización de pago en caja.
- Método de pago utilizado.

2. Pedidos y tiempos de preparación:

- Productos solicitados en cada pedido.
- Hora de inicio de preparación de cada producto.
- Hora de finalización de preparación de cada producto.

3. Selección y ocupación de mesas:

- Hora de inicio de ocupación de mesa.
- Hora de finalización de ocupación de mesa.

2.2. Instrumento de Medición

- **Aplicación web:** Se desarrolló un sistema de captura de datos basado en una página web en la cual todos los miembros del grupo participaron en las diferentes secciones que se muestran a continuación:

https://github.com/NicolasCacer/crepes_front

https://github.com/NicolasCacer/data_crepes

1. **Registro de atención y pago:** Captura los eventos de llegada, inicio de atención, pago y finalización de la atención en la caja.

2. **Registro de pedidos y tiempos de preparación:** Monitorea los tiempos de preparación de cada producto del pedido y cuenta los productos de cada pedido.
3. **Registro de ocupación de mesas:** Recopila los tiempos de ocupación y liberación de las mesas.

La figura 1 a continuación muestra el interfaz de la página principal de la página web.

Figura 1

Página Web Con sus Tres Secciones



Nota: Página elaborada por Nicolás Cáceres apoyada de Chat-GPT para su desarrollo.

- Todos los registros de tiempo son sincronizados mediante el reloj del servidor para garantizar precisión en la captura. Teniendo en cuenta hasta milisegundos.

2.3. Periodo de Observación

- Viernes a domingo de 12:00 p.m. a 3:00 p.m.
- Los viernes, el sistema inicia vacío (sin pedidos en cola ni clientes en fila).
- Los sábados y domingos, se registra el estado inicial desde las 11:30 a.m. antes de la toma formal de datos para caracterizar el estado inicial del sistema.

2.4. Procedimiento para toma de tiempos

1. **Proceso de atención y pago:**

- **Hora de arribo del cliente:** Se da click en la sección de la página donde dice Arribo a penas el cliente se pare en la fila de la caja. La página registra la hora y el día.
- **Hora de inicio de atención del pedido:** Se da click en la sección de la página donde dice Inicio Servicio Caja en el momento en que la persona comienza a ser atendido por la cajera.
- **Hora de inicio de atención del pago:** Se da click en la sección de la página donde dice Finalización Pedido en el momento en que la cajera le dice el total del pedido.
- **Hora de finalización de pago en la caja:** Se da click en la sección de la página donde dice Finalización Pago en el momento en que el cliente se retira de la caja.
- **Método de pago utilizado:** Se selecciona el método de pago a partir de una lista desplegable de la página que tiene las opciones: efectivo, tarjeta, bono de regalo, otro.

Los anteriores pasos se pueden ver en la figura 2 para mayor claridad. En el cuadro de observación se ponen algunos comentarios en caso de que el tiempo haya sido un error o algo atípico haya sucedido para poder tenerlo en cuenta en la limpieza y depuración de datos.

Figura 2

Sección 1 de la Página Web: Atención y Pago

Descripción	Arribo	Inicio Servicio Caja	Finalización Pedido	Finalización Pago	Método Pago	Observación	Acciones
<div>Eliminar</div> <input type="text"/>	17:42:13,824	-	-	-	Efectivo	<input type="text"/>	<div>Enviar</div> <div>Eliminar</div>

Agregar Registro

Nota: Página elaborada por Nicolás Cáceres apoyada de Chat-GPT para su desarrollo.

2. Pedidos y tiempos de preparación:

- **Productos solicitados en cada pedido:** Se genera una fila para cada pedido apenas entra en preparación. Según los tiempos que se toman en cada producto se sabe cuáles y cuantos productos hay en cada pedido.
- **Hora de inicio y fin de preparación de cada producto:** Se da click en el espacio al lado del tipo de producto donde inicia a tomar el tiempo. Para parar el tiempo se da click de nuevo. Y para otro producto de la misma categoría se vuelve a dar click para iniciarlo.

Es importante aclarar como se tomó el tiempo para cada tipo de producto.

- **Helados:** En el caso de los helados el tiempo de preparación se divide los vasos o conos y el artesano. Los conos o vasos de helado comienzan desde que el personal de preparación toma la galleta o vaso hasta que se entrega el helado al cliente; para los artesanos comienza desde el personal agarra el bote de helado hasta que entrega al cliente o guarda en la nevera.
- **Copas:** Se toma el tiempo desde que se toma el vaso hasta que se entrega al cliente.
- **Bebidas:** Se toma el tiempo desde que el personal de preparación toma el vaso o copa donde se va a servir la bebida hasta que se le hace entrega al cliente.
- **Gofres:** Se toma el tiempo desde que se vierte la mezcla para gofres en la waflera hasta que se entrega al cliente.
- **Crepes:** Se toma el tiempo desde que se vierte la mezcla para crepes en la plancha, hasta que se entrega al cliente.

En la figura 3 se ilustra lo que se dijo anteriormente. Se pueden tomar tantos tiempos por producto como se requieran. También se pueden agregar observaciones para tenerlas en cuenta a la hora de analizar datos.

Figura 3

Sección 2 de la Página Web: Pedidos y Tiempos de Preparación

Turno	Pedido	Observación	Acciones
<div>Turno</div>	HELADOS	Tres copas	<div>Enviar</div> <div>Eliminar</div>
	...		
	GOFRES		
	BEBIDAS		
	CREPES		


Nota: Página elaborada por Nicolás Cáceres apoyada de Chat-GPT para su desarrollo.

3. Selección y ocupación de mesas:

- **Hora de inicio de ocupación de mesa:** Si se toma la mesa se da click en la parte de la página donde dice ocupar mesa a penas el cliente se siente.
- **Hora de liberación de mesa:** Se da click en la parte de la página donde dice Liberar mesa cuando el cliente se haya parado de la mesa.

Figura 4

Sección 3 de la Página Web: Ocupación de Mesas

Descripción	Consumo	Ocupar Mesa	Liberar Mesa	Observación	Acciones
<input type="text"/>	<div>SI ▼</div>			<input type="text"/>	<div>Enviar  Eliminar</div>

[Agregar Registro](#)

Nota: Página elaborada por Nicolás Cáceres apoyada de Chat-GPT AI para su desarrollo.

3. Lineamientos para Depuración, Análisis y Validación

3.1. Depuración de Datos

Todos los datos se les realiza un histograma para ver su forma previamente y se le sacan algunas medidas muestrales como la media, la varianza, coeficiente de asimetría y coeficiente de curtosis. Esto con el objetivo de entender los datos, comprenderlos y realizar una revisión holística de posibles distribuciones que se ajusten. Posteriormente se realizan otras gráficas como la de violín, run sequence, p-p_plot y q-q plot.

En el caso del tiempo de pago, los datos se segmentaron según el método de pago: tarjeta, bono y efectivo. Dentro del grupo de tarjeta, se depuraron aquellas observaciones en las que se registró que el datáfono no estaba funcionando correctamente, para evitar distorsiones en el análisis. Por su parte, el tiempo de producto se dividió según el tipo de producto y luego

fue ajustado con base en observaciones puntuales realizadas durante la toma de datos, ya que en algunos casos se evidenciaron situaciones específicas que explicaban tiempos atípicos.

1. Eliminación de registros inconsistentes:

- Datos que se tomaron mal y tienen algún tipo de observación en la que se dice lo que ocurrió. Como tiempos enviados por accidente.
- Casos en los que el flujo lógico de eventos esté alterado (por ejemplo, un pedido procesado antes de su pago).

2. Realizar desplazamientos y Ajustes (Normalizar):

Este procedimiento se aplica a las distribuciones discretas presentes en el proyecto: la cantidad de productos, el tipo de producto, el medio de pago y si ocupa o no mesa. Para cada una de estas variables se verifica si es necesario realizar un desplazamiento a cero antes del análisis, ya que algunas distribuciones discretas, como las binomiales, requieren que los valores comiencen desde 0 para su correcta interpretación. Este proceso se realiza utilizando Python, específicamente con la librería NumPy.

Sin embargo, solo fue necesario aplicar este desplazamiento a las variables que presentaban un comportamiento binomial en el medio de pago (cuando se analiza la categoría "tarjeta") se comportó como una binomial, ya que permitía identificar dos estados: si el datáfono estaba funcionando o si estaba dañado. Por ello, se generaron dos distribuciones separadas para esa categoría.

De forma similar, en el tipo de producto, específicamente en los gofres, se identificó una distribución binomial que distinguía entre aquellos casos en los que la masa comenzó a prepararse desde cero y aquellos en los que se tomaba directamente de la waflera, dado su comportamiento binomial, ambas variables fueron desplazadas para iniciar desde 0, con el fin de garantizar una correcta interpretación de los datos. Las demás variables, al presentar distribuciones empíricas, no requirieron este tipo de ajuste. Las demás variables, al presentar distribuciones empíricas, no requirieron desplazamiento, ya que no tienen una estructura que exija iniciar en cero para su análisis.

3. Identificación de valores atípicos:

Esto se debe realizar especialmente en distribuciones continuas, utilizando herramientas gráficas como histogramas, diagramas de violín, boxplots y gráficos QQ (quantile-quantile), que permiten visualizar posibles desviaciones significativas respecto al comportamiento general de los datos.

4. Tratamiento de valores atípicos:

En la mayoría de los casos, los valores atípicos fueron identificados y reemplazados por la mediana general de la variable correspondiente. Sin embargo, en el caso del tiempo de pago, se optó por un tratamiento más específico: se calcularon las medianas de manera separada para cada método de pago (tarjeta, bono y efectivo), y los valores atípicos fueron reemplazados por la mediana correspondiente a su categoría. Para validar este ajuste, se analizaron los histogramas y otras gráficas asociadas. En los casos en los que las distribuciones seguían mostrando inconsistencias o no se ajustaban adecuadamente, se procedió a segmentar aún más los datos, dividiéndolos por tipo de producto, de modo que las medianas utilizadas para el reemplazo reflejaran con mayor precisión el comportamiento particular de cada subgrupo.

5. Identificación y Tratamientos de sesgos:

Inicialmente, se evaluó cuál era la distribución que mejor se ajustaba a cada conjunto de datos. En los casos en que todas las distribuciones fueron rechazadas, se recurrió al análisis visual mediante histogramas. Esto permitió identificar posibles sesgos o comportamientos atípicos en los datos. Cuando se detectaron patrones inconsistentes, se procedió a segmentar los datos por clases de tiempo, con el fin de analizar los sesgos de manera independiente y estimar sus distribuciones por separado.

En cuanto a los valores atípicos, se observó que algunos no representaban errores puntuales, sino que en realidad correspondían a una distribución distinta, lo que llevó a la identificación de distribuciones bimodales. Esta diferenciación fue clave para no eliminar datos válidos que, aunque alejados de la media, reflejaban un comportamiento real perteneciente a otra subpoblación dentro del conjunto.

3.2. Análisis de Datos

El análisis de datos y asignación de distribuciones se realizará por medio de Python y las librerías de matplotlib, pyplot, sklearn, pandas, seaborn, scipy.stats, numpy, statsmodel, scipy.stats.

1. Verificación de estacionalidad de las variables

Se debe revisar la estacionalidad de todas las distribuciones tanto discretas como continuas para ver si alguna variable cambia dependiendo del día de la semana e incluso dependiendo de la hora. Para saber si se deben separar para realizar un modelo más preciso. Esto se realizará mediante los gráficos de violín, ANOVA y Tukey. Para probar estas diferencias y ver si cumplen o no con la propiedad estacionaria.

2. Separación de Distribuciones Multivariadas

De la mano de la estacionalidad se debe tener en cuenta que si la variable es multivariada puede ser por la estacionalidad o por otros motivos como el método de

pago la cantidad de personas por grupo entre otros factores por lo que se debe verificar esto para saber si se debe separar y con que criterio se separa. Para saber como separarlos se usarán dos límites de separación y luego se ajustarán los datos para cada límite y el que obtenga el valor p más alto se definirá como límite de separación de distribución. También se busca comprender la causa de que sea multivariada por lo que se deben analizar la correlación entre otras variables esto por medio de pruebas de correlación como la de Pearson.

3. Ajustar Parámetros

Para cada variable se realizarán pruebas de bondad de ajuste (Kolmogorov-Smirnov, Chi-cuadrado) para las siguientes distribuciones en caso de ser discretas: Poisson, uniforme discreta, binomial, binomial negativa y empíricas. Y se realizará el ajuste de parámetros teniendo en cuenta el modelo conceptual, la facilidad de modelado de las funciones, así como su valor p para ver que tanto se ajustan.

En el caso de las variables continuas, se realizarán pruebas de bondad de ajuste para las siguientes distribuciones: Normal, Gamma, Beta, Lognormal, Exponential, Triangular, Uniforme, Weibull. Se seleccionan estas distribuciones porque son las más conocidas lo que hace que estén bien documentadas, que sean fáciles de modelar en los software, no tienen tantos parámetros por lo que se pierden menos grados de libertad y estas distribuciones no requieren tanto recurso computacional como otras más complejas.

3.3. Validación del Modelo

1. Comparación con datos reales:

- Se comparan los resultados del modelo con los datos recolectados para evaluar su representatividad. Esto se realiza con los datos de cuánto tiempo el cliente esperaba en la fila, la cantidad de personas en fila. La cantidad de pedidos en cola. Estos datos se calculan a partir de la misma toma de datos porque se pueden calcular por transformación de datos ya que cada registro tiene su id.

2. Pruebas estadísticas:

- Uso de intervalos de confianza para medir la precisión del modelo.
- Uso del P-valor en pruebas de bondad de ajuste y chi-cuadrado, para saber si se ajustan o no se ajustan los datos a la distribución que se propone.