

Rendimiento Web de Vanguard: Un Estudio Basado en Datos con Pruebas A/B

Analizaremos los datos del **rendimiento** web de Vanguard, una empresa líder en el sector de gestión de **activos**, para obtener **insights** valiosos que impulsen su **éxito** digital.

Por Irene Sifre y Adrián Lardiés



Introducción

Pregunta Principal

Misión

- Proporcionar a Vanguard análisis de datos **detallados** sobre la renovación de su web
- Asegurar un **impacto** positivo en la experiencia del cliente.

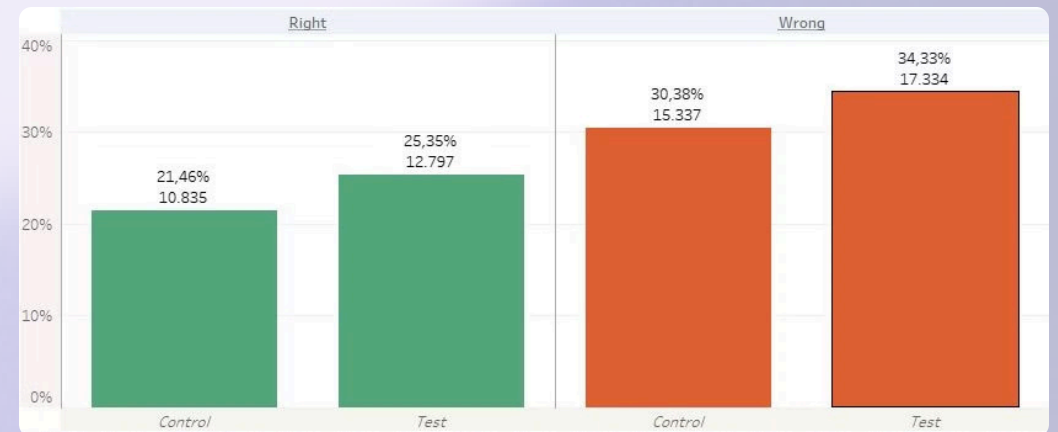
Presencia

- Vanguard, **líder** global en el sector financiero, reconocida por su **innovación** y confianza.

Retos

- **Segmentar** a los clientes más allá de los grupos de control y test para generar **insights** precisos.
- Ofrecer a Vanguard máxima **fiabilidad** y una visión más amplia del **comportamiento** de sus usuarios.

- ¿Ha **mejorado** la nueva interfaz de usuario las **tasas** de éxito general en **comparación** con la versión anterior?



Limpieza y Fusión de los Datos

Carga

1

- `df_final_demo`: Información **demográfica** de los clientes.
- `df_final_experiment_clients`: Información de los clientes que **participaron** en el experimento.
- `df_pt`: Datos del proceso de **navegación** del cliente (de dos archivos combinados).

Limpieza

2

- **Eliminación de duplicados**: Se eliminaron registros duplicados en todos los conjuntos de datos.
- **Manejo de valores faltantes**: Se eliminaron los registros con valores nulos en columnas clave.

Filtrado

3

- Se identificaron los clientes **comunes** entre los tres conjuntos de datos utilizando el `client_id`.
- Se filtraron los datos para mantener solo los clientes presentes en **todos** los conjuntos de datos.

Conversión

4

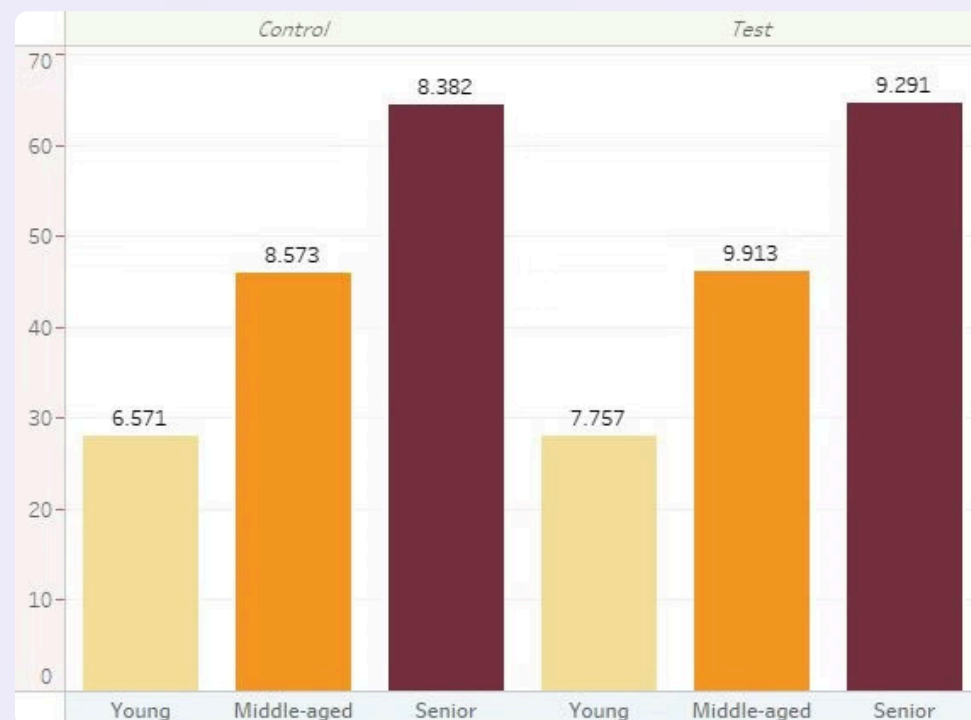
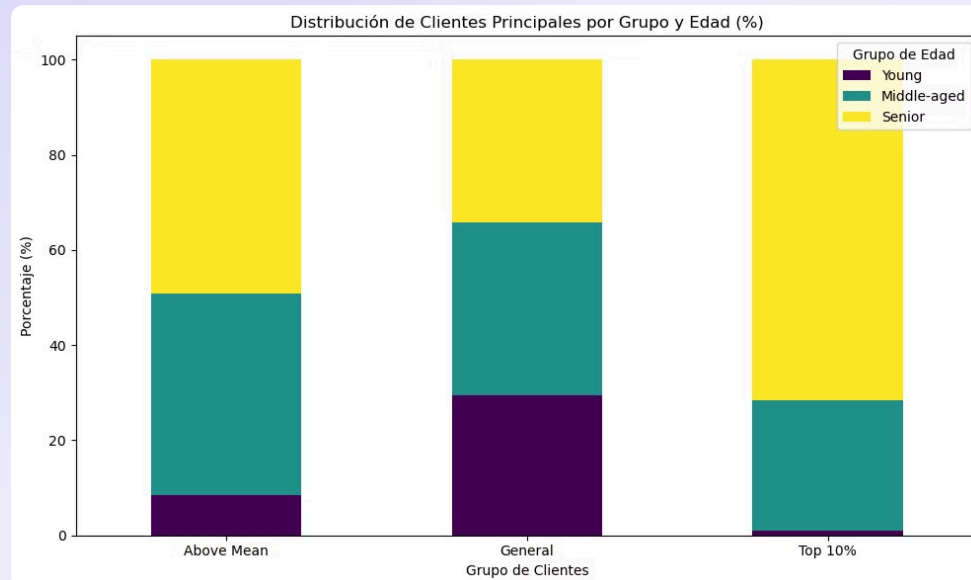
- Se **convirtieron** las columnas numéricas (como años de antigüedad, número de cuentas, llamadas, logins) a formato `int`.
- Se convirtieron las fechas a formato `datetime` para un procesamiento adecuado de los **tiempos** de navegación.

Renombrado y Fusión

5

- Se **renombraron** las columnas en `df_final_demo` y `df_pt` para consistencia (ej. `'clnt_tenure_yr'` a `'tenure_years'`, `'process_step'` a `'step'`).
- Se **fusionaron** los datos demográficos (`df_final_demo`) con los datos del experimento (`df_final_experiment_clients`) en un único DataFrame (`df_final`).
- Finalmente, se fusionó `df_final` con los datos de navegación (`df_pt`) para **construir** el DataFrame final (`df_vanguard`).

Perfiles de Clientes

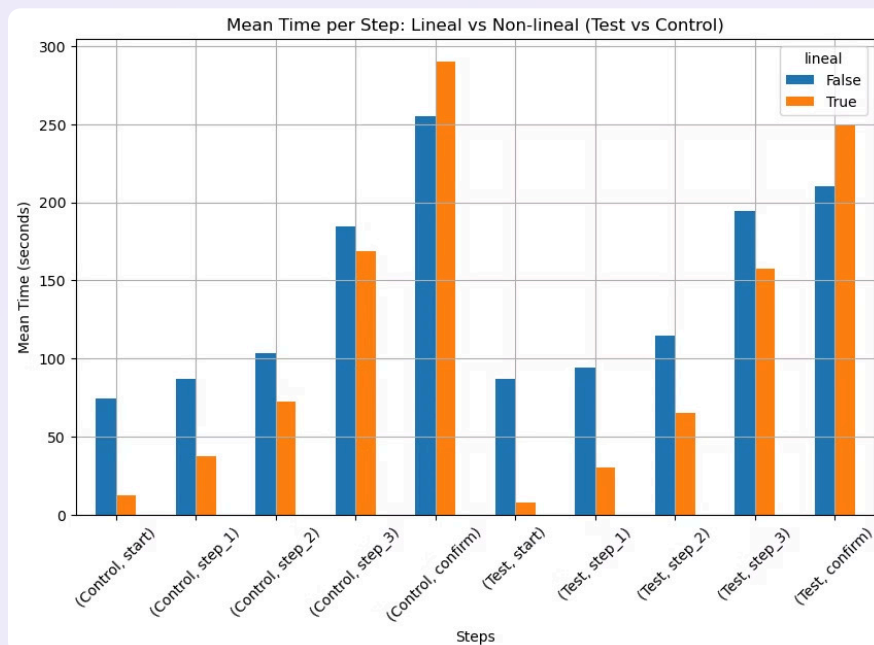
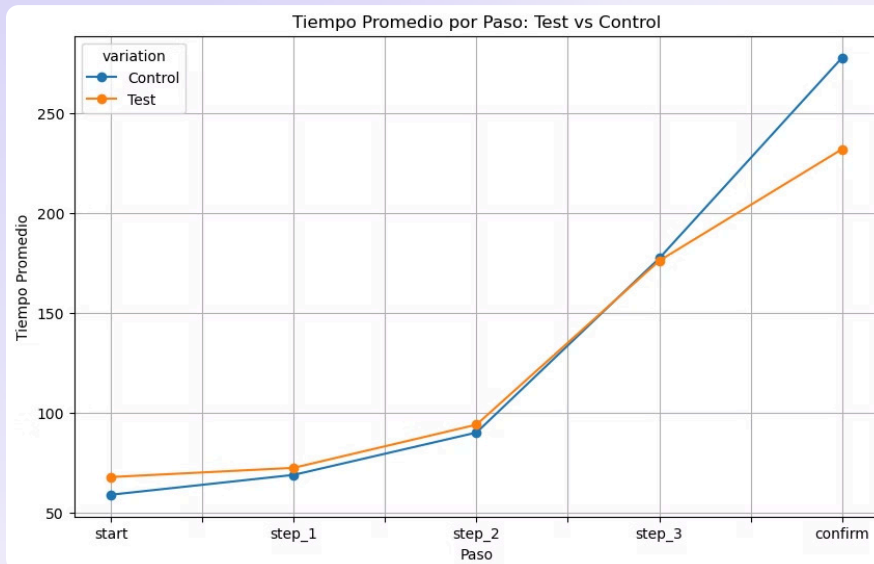


- **Variables** como edad, balance, número de cuentas, actividad en la plataforma (llamadas y logins), y tiempo de permanencia.
- Los clientes han sido **segmentados** en tres grupos principales: **General**, **Above Mean** y **Top 10%** de acuerdo a las variables previas.
- Además, en tres grupos de **edad**: **Young** (avg. 28), **Middle-aged** (46), y **Senior** (65). Cada grupo muestra comportamientos distintos, con **balances** promedio que van desde 34771 \$ (Low) hasta 351958 \$ (High).

Insights

- El grupo que analiza la nueva versión supera de media en **1000 clientes** por categoría a la versión antigua, lo que podría influir en los resultados del análisis de la web.
- Las personas que utilizan el sitio **correctamente** tienen niveles de **actividad** más **bajos**, podríamos pensar que llaman menos porque están mejor **adaptadas** o son más **resolutivas**.
- Al mismo tiempo, las personas que participan en la versión **antigua** del estudio tienen un **menor** nivel de **actividad**.

Huellas Digitales

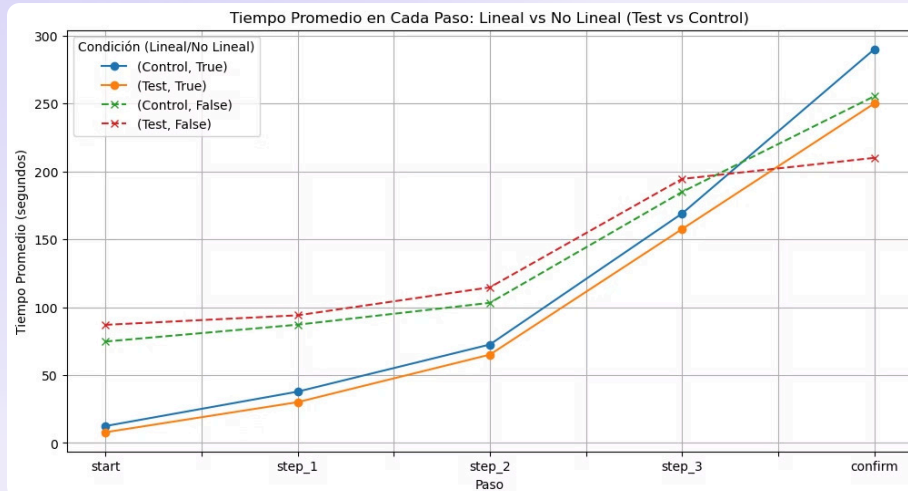


- Se registraron las interacciones de los clientes con la web a través de un **sistema de pasos** (start, step_1, step_2, step_3, confirm), analizando el **orden**, **repeticiones** y **tiempos** de cada paso.
- Se creó la variable "**lineal**" para identificar si los clientes completaron el proceso **sin errores**, y "**non_linear**" para los que no siguieron el proceso de forma correcta.

Insights

- Los usuarios **lineales** completan el proceso con **tiempos más bajos** en todos los pasos (excepto **Confirm**), especialmente en la versión de **prueba** (Test).
- Los usuarios **no lineales** muestran tiempos significativamente mayores, y los **cuellos de botella** más comunes aparecen en **step_3** y **confirm**.
- En el grupo **Test**, los tiempos **mejoraron** significativamente en los pasos **intermedios** para los usuarios lineales, lo que sugiere un impacto **positivo** de la nueva interfaz.

Lista de Experimentos



- Se evaluaron tres aspectos **clave** para medir el rendimiento entre los grupos de Test y Control, y entre usuarios lineales y no lineales: **tasa de finalización**, tasa de **errores** por **pasos**, y **tiempo** medio en cada paso.

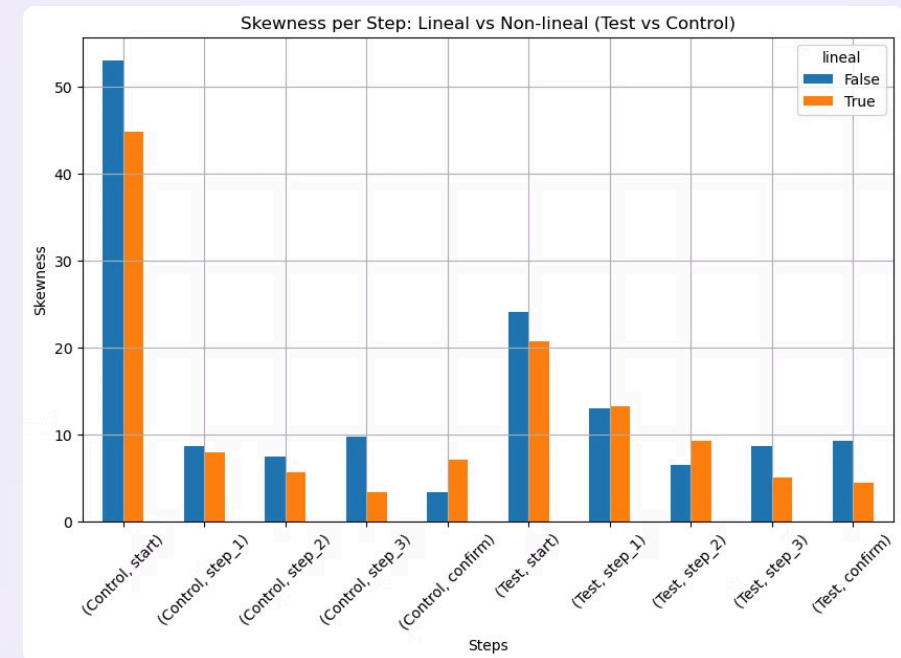
- Insights

- La tasa de **finalización** fue **mayor** en el grupo de **Test** (25.35% lineales) en comparación con el grupo de Control (21.46% lineales).
- La tasa de **errores** en **Start** es más **alta** en **Test**, podríamos interpretar que se debe al hecho de que los usuarios estaban **habituados** al uso clásico de la web.
- Step 1** es notablemente **deficiente** en la **nueva** versión al mismo tiempo.
- Focalizar la experiencia **UX** en Start, Step 1 y Step 2 en la **nueva** versión.

Skweness y Kurtosis por Steps y Linealidad

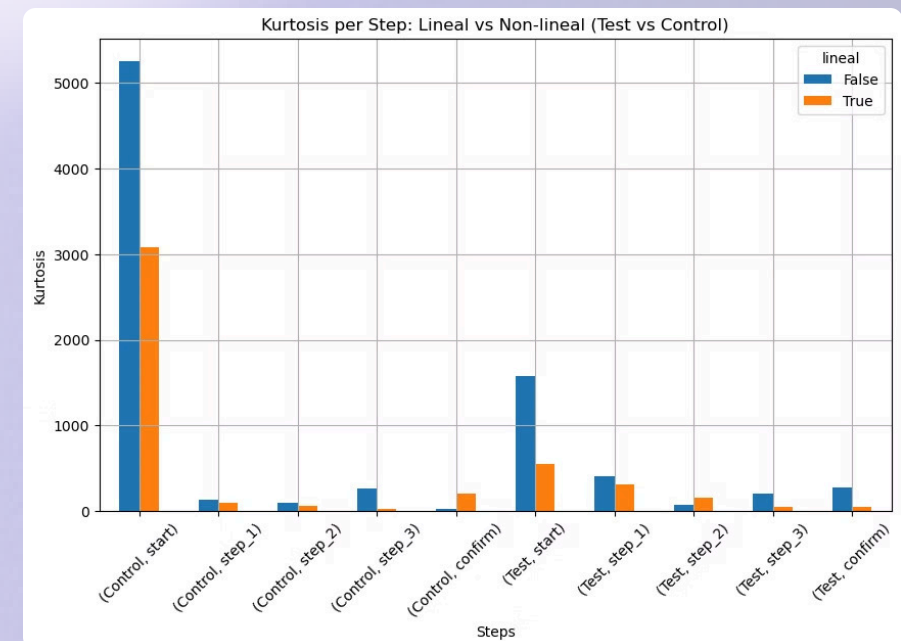
- **Skweness**

- Permite identificar qué tan sesgados están los tiempos hacia un lado de la media.
- Valores altos en pasos como **start** (por ejemplo, 53.05 en **Control-False**) indica que la mayoría de los usuarios pasan poco tiempo, pero algunos tienen tiempos excesivamente largos en estos pasos.



- **Kurtosis**

- Indica la concentración de los tiempos alrededor de la media. Valores extremadamente altos, como 5255.29 en **Control-False-start**, muestran una gran variabilidad en los tiempos, con algunos usuarios pasando tiempos muy largos en los pasos iniciales, lo que puede ser indicativo de atascos o errores.
- En general, los usuarios **lineales (True)** tienen una kurtosis baja, reforzando nuestra división para el análisis.



Pruebas de Hipótesis y Estadísticas

Planteamiento

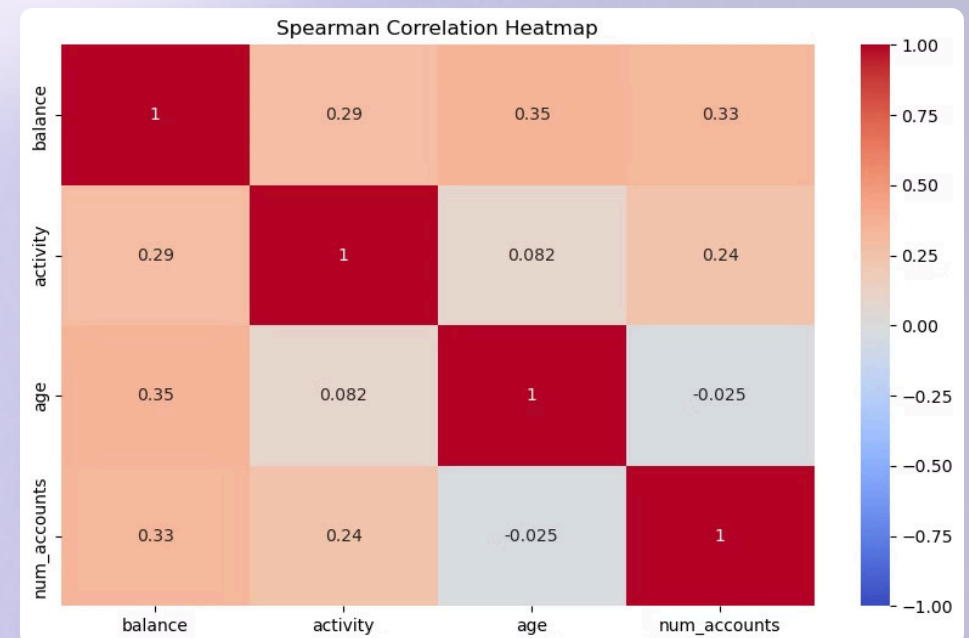
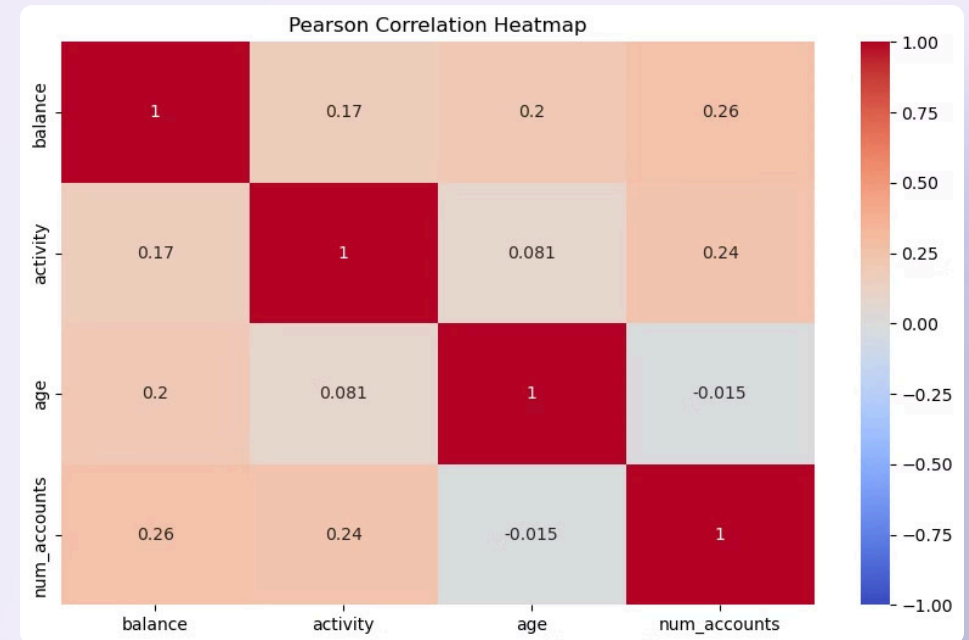
- **Hipótesis nula (H0):** No hay diferencias significativas entre la versión Test y Control en términos de tasas de finalización, tiempos de finalización, tasas de error, etc.
- **Hipótesis alternativa (H1):** Hay una diferencia significativa entre la versión Test y Control en al menos una de las métricas.

Pruebas

- **Prueba de proporciones (Prueba Z):**
 - Comparar las tasas de finalización entre los grupos Test y Control.
 - Resultados diferencia estadísticamente significativa ($p < 0.05 \rightarrow 0.00154$), indica que el comportamiento de los usuarios difiere entre ambas versiones.
- **T-test / Mann-Whitney U Test:**
 - **T-test** para comparar los tiempos medios de finalización entre Test y Control.
 - Diferencia significativa en los tiempos de finalización, con $p < 0.05 \rightarrow 1.145e-65$.
 - Dado que los datos no seguían una distribución normal (como se evidenció en las pruebas de normalidad), se utilizó la **prueba Mann-Whitney** como alternativa no paramétrica, y también mostró una diferencia significativa en los tiempos.
- **Chi-square test:** para comparar la distribución de errores y finalización entre Test y Control.
 - Resultados significativos ($6.26e-31$), diferencia en la forma en que los usuarios completan los pasos en ambas versiones.
- **Cramér's V:** Aunque se detectó una diferencia significativa en el comportamiento de los usuarios entre Test y Control, el valor de Cramér's V fue bajo (0.0205).
 - Indica que la asociación entre las versiones de la web y el comportamiento de los usuarios es débil.

Correlación entre Variables

- **Pearson:** para evaluar la relación lineal entre variables continuas como balance, actividad, edad y número de cuentas.
 - Se encontraron correlaciones positivas bajas entre **balance y actividad** (0.171) y correlaciones moderadas entre **balance y número de cuentas** (0.264), lo que indica una relación entre tener más cuentas y un balance más alto.
- **Spearman:** dado que algunas relaciones pueden no ser lineales, se aplicó Spearman, que mostró correlaciones más fuertes entre **balance y actividad** (0.290) y **balance y número de cuentas** (0.333), sugiriendo que los clientes más activos y con más cuentas tienden a tener mayores balances.



TLC (Teorema del Límite Central) y Normalización

- Aunque el **TLC** permite asumir que, con muestras grandes, las medidas de los parámetros se distribuyen normalmente, las pruebas de normalidad realizadas (como **Shapiro-Wilk** y **Kolmogorov-Smirnov**) indicaron que los datos no seguían una distribución normal.
 - Esto motivó la elección de la **prueba Mann-Whitney** en lugar del T-test.
- Se intentó **normalizar** los tiempos mediante transformaciones **logarítmica**, **raíz cuadrada** y **Box-Cox**, pero ninguna logró normalizar completamente los datos.

