

Análisis de Datos en la Aviación: Dominando Variables Categóricas

De la tabla de contingencia a la visualización
estratégica de flotas y tripulaciones.

[SYSTEM_STATUS: **ONLINE** ▶ | DATA_SET: **10,000_RECORDS** 📱 | MODULE: 📊 **STAT_01**]

El Tablero de Vuelo

```
CMD_093: STS_DLY: TYP_M: ALT_340
ACFT_737: RTE_MAD_BCN: OPR_IBE
FLT_505: STS_ON_TIME: CREW_COM
FLT_505: STS_ON_TIME: CREW_COM
DATA_STREAM: CHAOS_LEVEL_HIGH: VAR_QUAL
```



EXP_OOR: 72%
EXP_REL: 38%

EXP_OOR: 72%
EXP_REL: 28%
TOTAL_CREW: 1458

TOTAL_CREW: 1458

Tripulación

Antecedentes del Piloto
(Comercial vs. Militar).
¿Cómo se distribuye la
experiencia?

```
CMD_093: STS_DLY: TYP_M: ALT_340
CMD_092: STS_DLY: TYP_M: ALT_340
ACFT_737: RTE_MAD_BCN: OPR_IBE
FLT_505: STS_ON_TIME: CREW_COM
FLT_505: STS_ON_TIME: CREW_COM
DATA_STREAM: CHAOS_LEVEL_HIGH: VAR_QUAL
```

```
CMD_093: STS_DLY: TYP_M: ALT_340
ACFT_737: RTE_MAD_BCN: OPR_IBE
FLT_505: STS_ON_TIME: CREW_COM
FLT_505: STS_ON_TIME: CREW_COM
DATA_STREAM: CHAOS_LEVEL_HIGH: VAR_QUAL
```



ETS_DLY: 19%
AVIATION ALERT: 19%

CERT_MULTI: 60%
CERT_SINGLE: 20%
CERT_ROTATOR: 15%

FLEET_SIZE: 320

Aeronaves

Certificación de Vuelo
(Monomotor, Multimotor,
Ala Rotatoria).
¿Cuál es la composición
de la flota?

```
CMD_093: STS_DLY: TYP_M: ALT_340
CMD_093: STS_DLY: TYP_M: ALT_340
ACFT_737: RTE_MAD_BCN: OPR_IBE
FLT_505: STS_ON_TIME: CREW_COM
FLT_505: STS_ON_TIME: CREW_COM
DATA_STREAM: CHAOS_LEVEL_HIGH: VAR_QUAL
```

```
CMD_093: STS_DLY: TYP_M: ALT_340
ACFT_737: RTE_MAD_BCN: OPR_IBE
FLT_505: STS_ON_TIME: CREW_COM
FLT_505: STS_ON_TIME: CREW_COM
DATA_STREAM: CHAOS_LEVEL_HIGH: VAR_QUAL
```



AVIATION ALERT: 19%

STS_ON_TIME: 81%
STS_DLY: 19%
AVG_DLY: 32 MIN



Operaciones

Estado del Vuelo (A
Tiempo vs. Retrasado).
¿Dónde ocurren las
disrupciones?

```
CMD_093: STS_DLY: TYP_M: ALT_340
CMD_093: STS_DLY: TYP_M: ALT_340
ACFT_737: RTE_MAD_BCN: OPR_IBE
FLT_505: STS_ON_TIME: CREW_COM
FLT_505: STS_ON_TIME: CREW_COM
DATA_STREAM: CHAOS_LEVEL_HIGH: VAR_QUAL
```

Los datos categóricos exigen **herramientas** específicas. Esta presentación es su **manual de vuelo** para estructurar, **visualizar** y cruzar variables **cualitativas** con precisión matemática.

Tablas de Contingencia: Estructurando el Caos

Estructura de la Flota y Tripulación

Certificación:	Monomotor	Multimotor	Ala Rotatoria	Total
Piloto Comercial	3496	3839	1170	8505
Piloto Militar	362	950	183	1495
Total	3858	4789	1353	10000

Cada celda representa la intersección de dos variables.

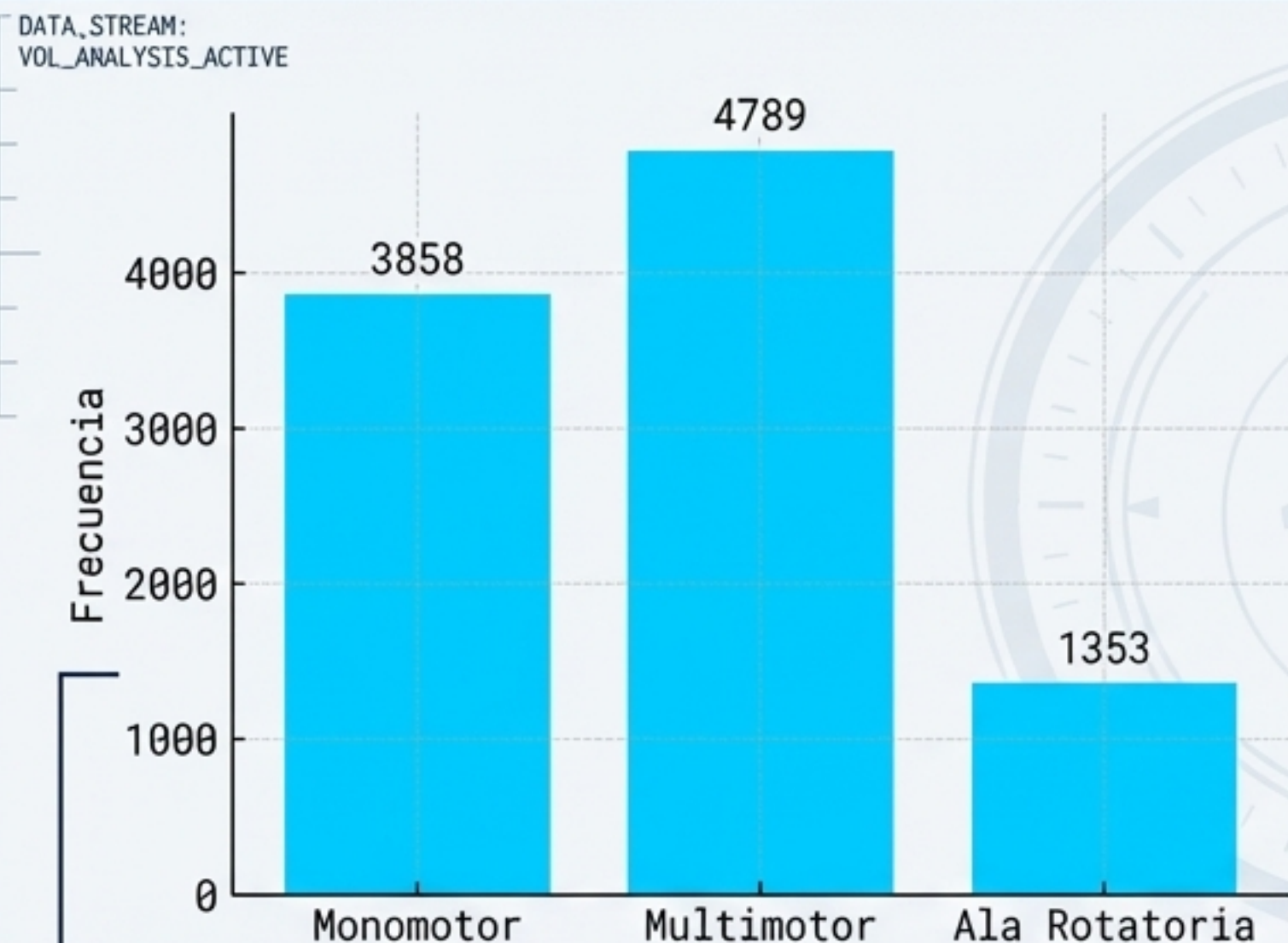
Ejemplo: Existen exactamente 3,496 casos en nuestra muestra de 10,000 donde un piloto comercial está certificado para aeronaves monomotor.

STS_GR: 319%
AVG_DLY: 355H

OND_GRP: STS_DLY: TYP_MG
ACT_SBY: STS_MFDU_KOCH: S

Frecuencias vs. Proporciones: Decodificando el Volumen

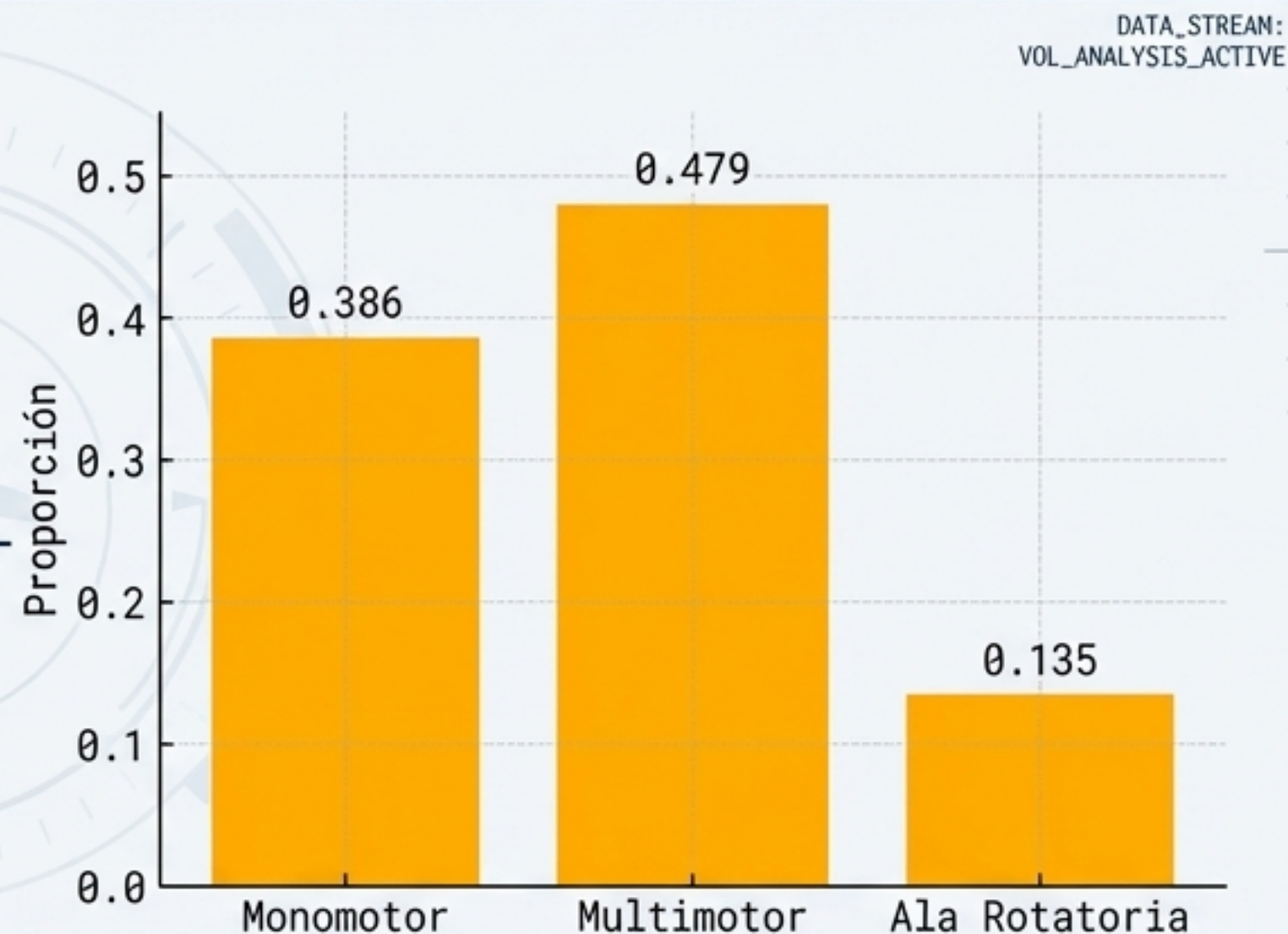
Frecuencia Absoluta



Los conteos brutos muestran el volumen total. Útil para logística y dimensionamiento.

DATA_STREAM:
VOL_ANALYSIS_ACTIVE

Proporción Relativa



Las proporciones (conteo dividido por 10,000) revelan la estructura subyacente. El 47.9% de las certificaciones son multimotor, independientemente del volumen total.

Proporciones de Fila vs. Columna: La Pregunta Define la Matemática

DATA_STREAM:
VOL_ANALYSIS_ACTIVE

DATA_STREAM:
VOL_ANALYSIS_ACTIVE

Analizando 3,921 vuelos.
Nacionales (1,195) vs. Internacionales (2,726).
Estado: Retrasado (367) vs. A Tiempo (3,554).

Path A

Proporción de Columna: ¿Qué tipo de vuelo es más propenso a retrasos?

$$\frac{\text{Vuelos Nac. Retrasados (209)}}{\text{Total de Nacionales (1195)}} = 17.5\%$$

$$\frac{\text{Vuelos Int. Retrasados (158)}}{\text{Total de Int. (2726)}} = 5.8\%$$

Usamos la variable explicativa (Tipo de Vuelo) como denominador. Revela que los vuelos nacionales tienen un riesgo de retraso drásticamente mayor.

DATA_STREAM:
VOL_ANALYSIS_ACTIVE

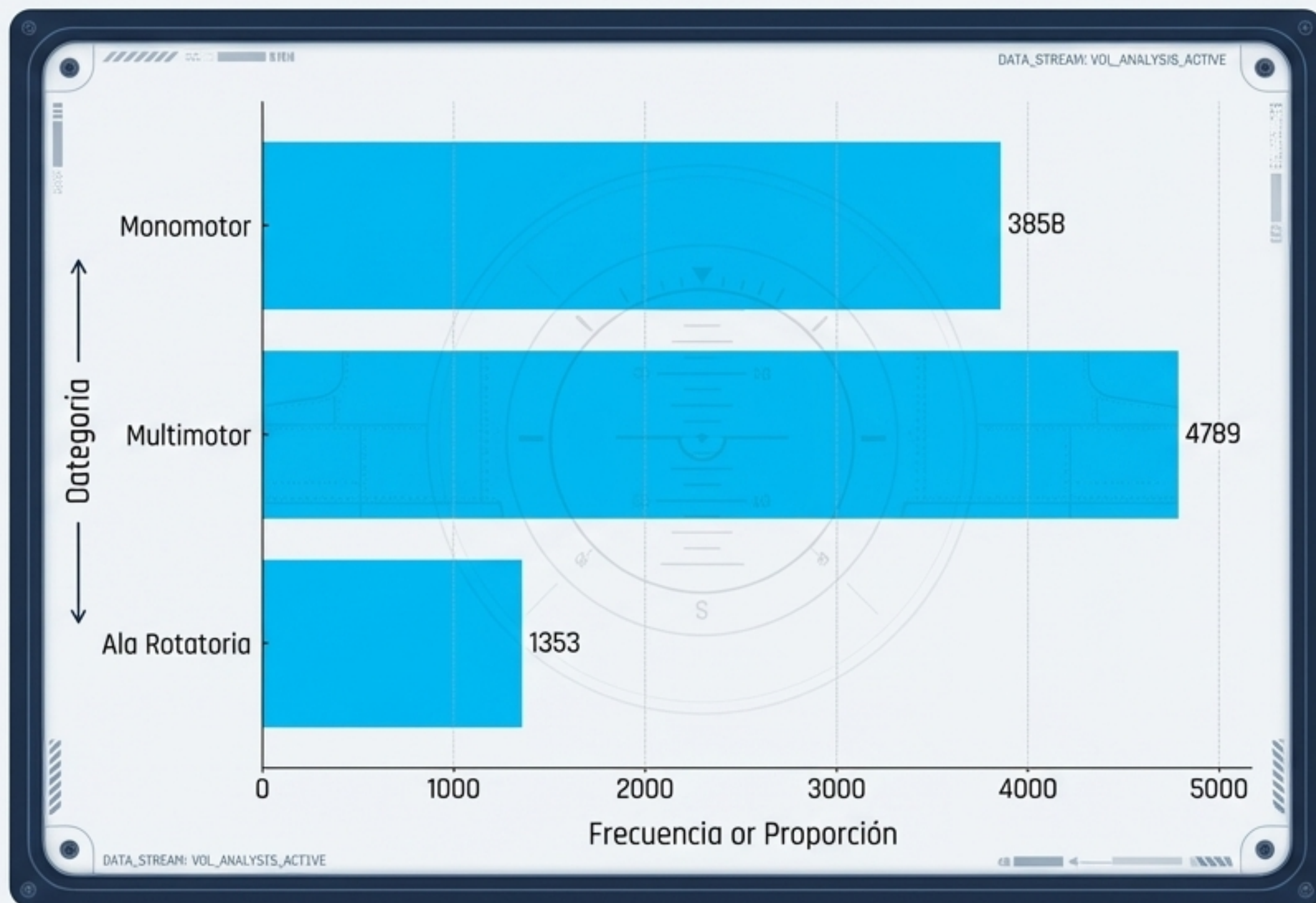
Path B

Proporción de Fila: ¿De dónde provienen nuestros retrasos totales?

$$\frac{\text{Nacionales Retrasados (209)}}{\text{Total de Retrasos (367)}} = 56.9\%$$

Cambiar el denominador cambia el enfoque operativo.

Visualizando el Terreno: El Gráfico de Barras Simple



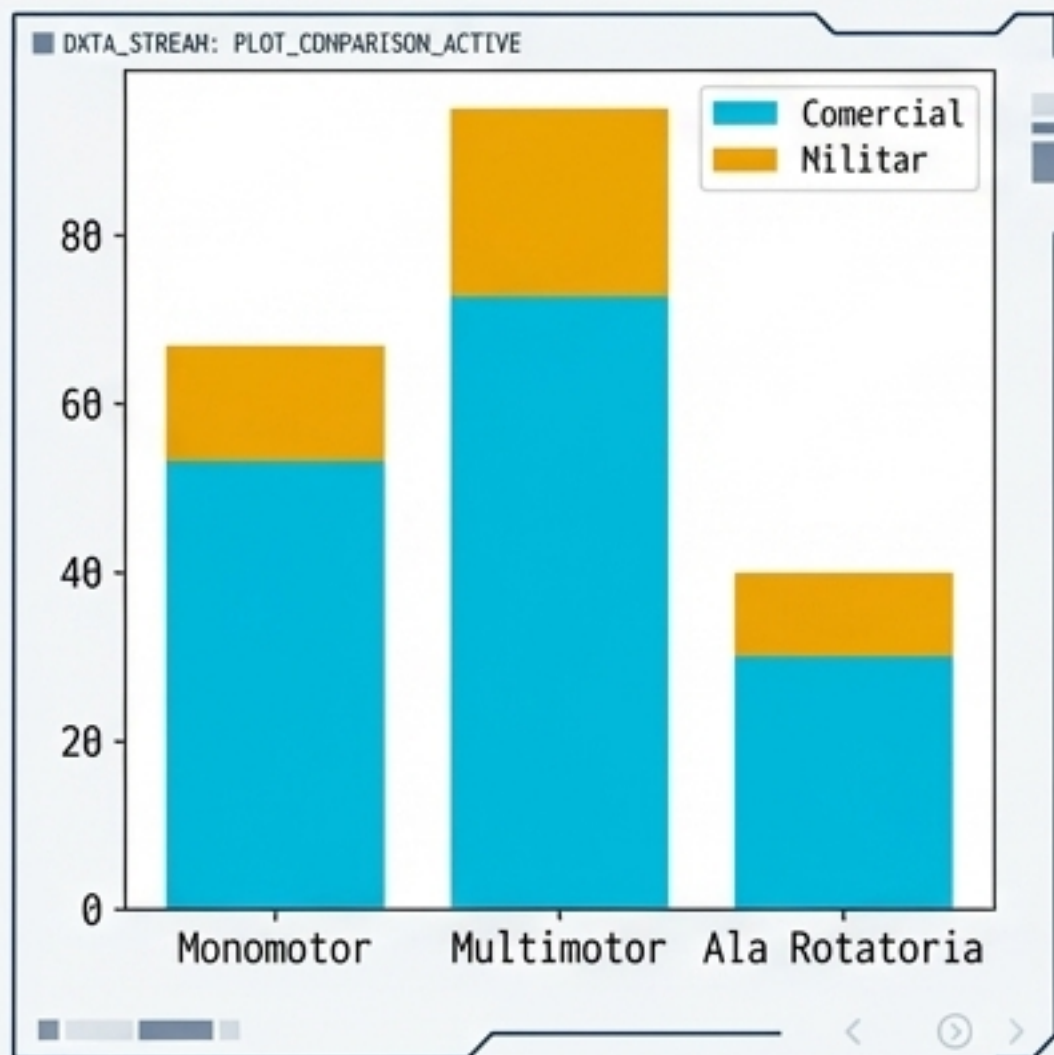
Regla de Oro: La forma más eficiente de comparar una única variable categórica.

Cognición Rápida: El ojo humano compara longitudes con extrema precisión. En fracciones de segundo, es evidente que Multimotor domina la flota.

Flexibilidad: El Eje X puede representar **Frecuencia** (conteos brutos) o **Proporción** (porcentajes), dependiendo del objetivo del análisis.

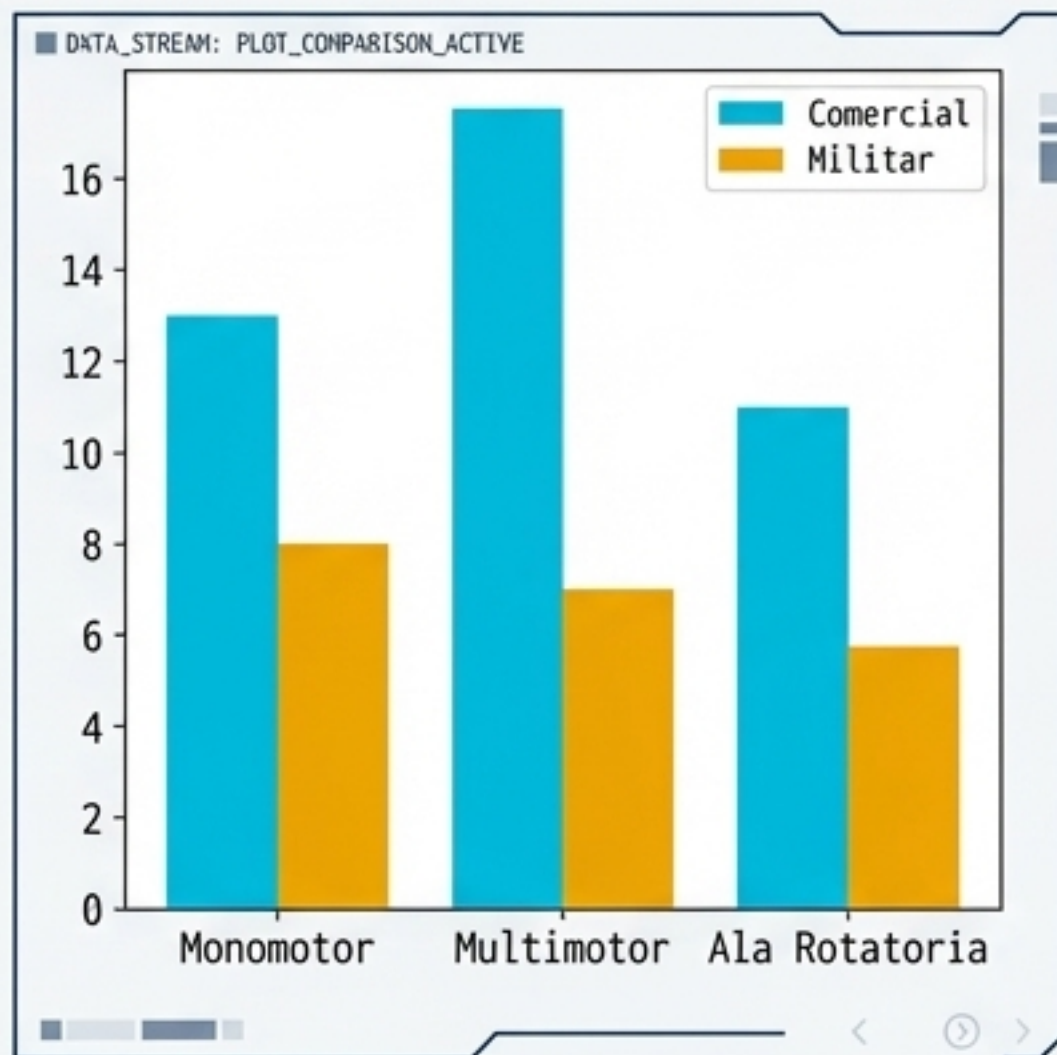
Gráficos de Barras para Dos Variables: Tres Perspectivas

Gráfico Apilado (Stacked)



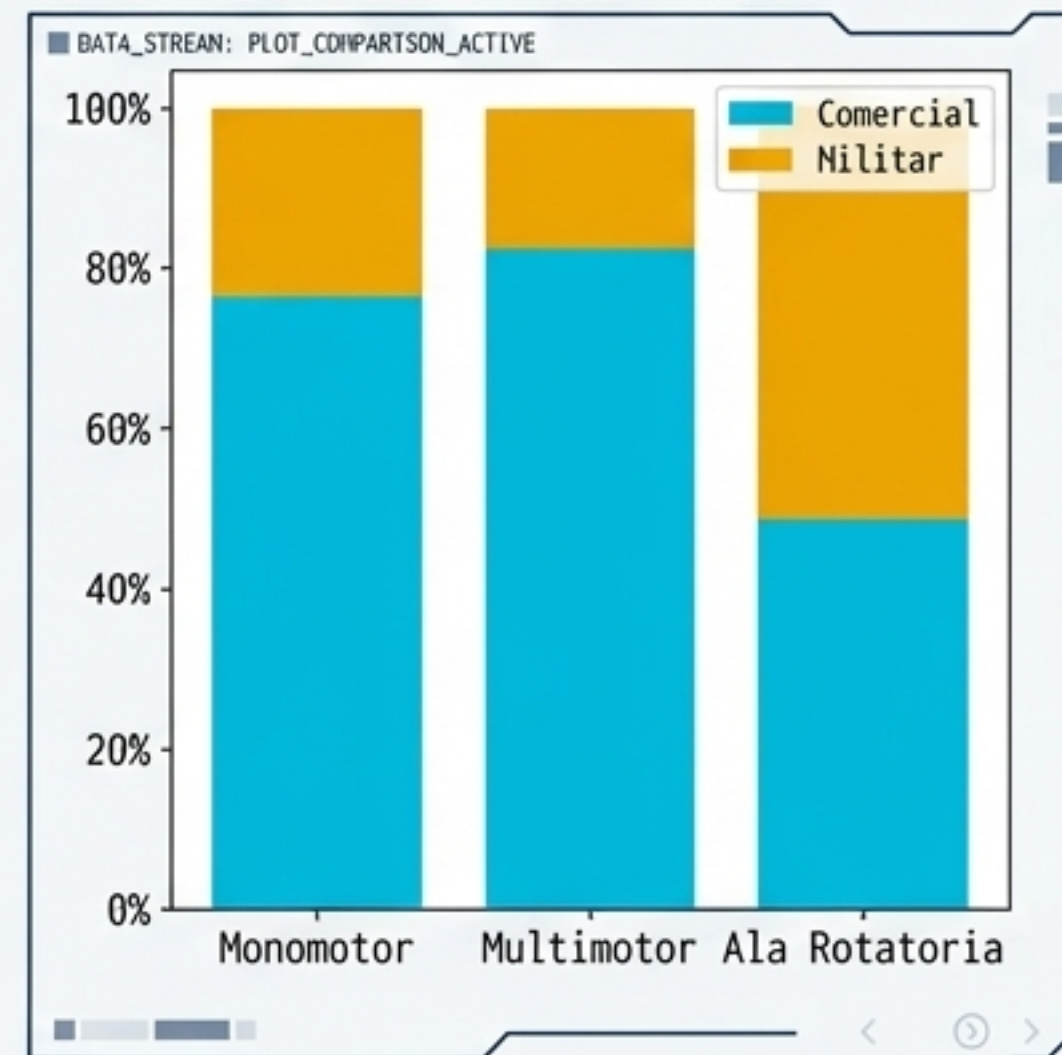
Agrupar por variable primaria, subdivide por la secundaria.

Lado a Lado (Side-by-side)





Agnóstico. Permite ver claramente el volumen de las 6 combinaciones.

Apilado Estandarizado (Standardized)



Elimina la disparidad de volumen para aislar puramente la asociación entre variables.

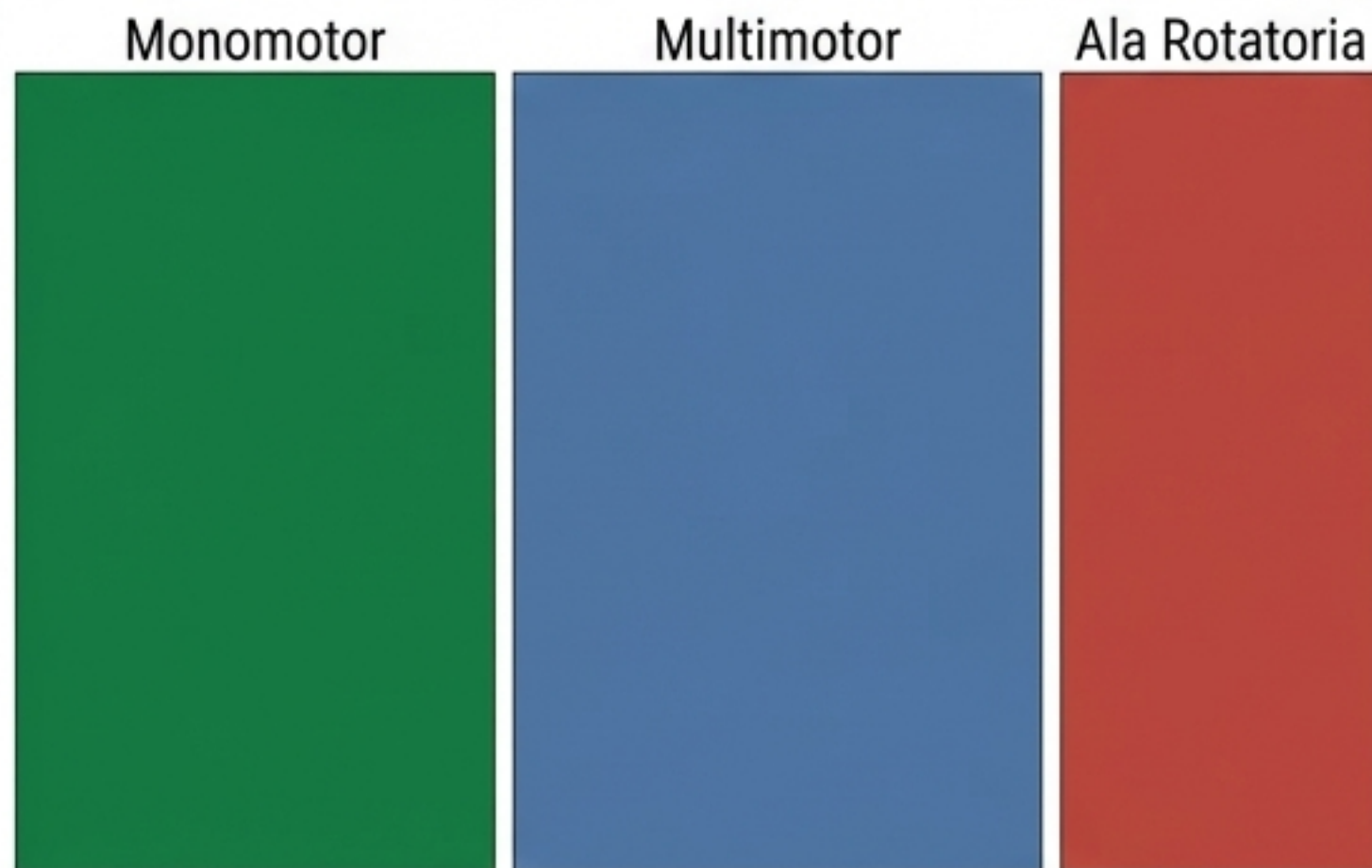
Matriz de Decisión Analítica: ¿Cuál Gráfico Usar?

Variante	Objetivo Analítico	Ventaja Principal 	Punto Ciego Operativo 
Apilado (Stacked)	Mostrar volumen total mientras se asume una variable explicativa.	Excelente visión del tamaño total de cada categoría primaria.	Difícil comparar los segmentos superiores entre diferentes barras.
Lado a Lado (Side-by-side)	Comparación directa e independiente de todas las subcategorías.	Fácil lectura del número exacto de casos por cada intersección.	Requiere mucho espacio horizontal; puede verse apretado y dificulta ver asociaciones si hay desequilibrios severos.
Apilado Estandarizado (Standardized)	Descubrir asociaciones entre variables altamente desequilibradas.	Revela inmediatamente si las proporciones cambian entre grupos.	Se pierde absolutamente toda noción de la cantidad de casos totales.

El Gráfico de Mosaico: Volumen y Proporción en Simultáneo

Paso 1: Eje Primario

DATA_STREAM: PRIMARY_AXIS_BUILD

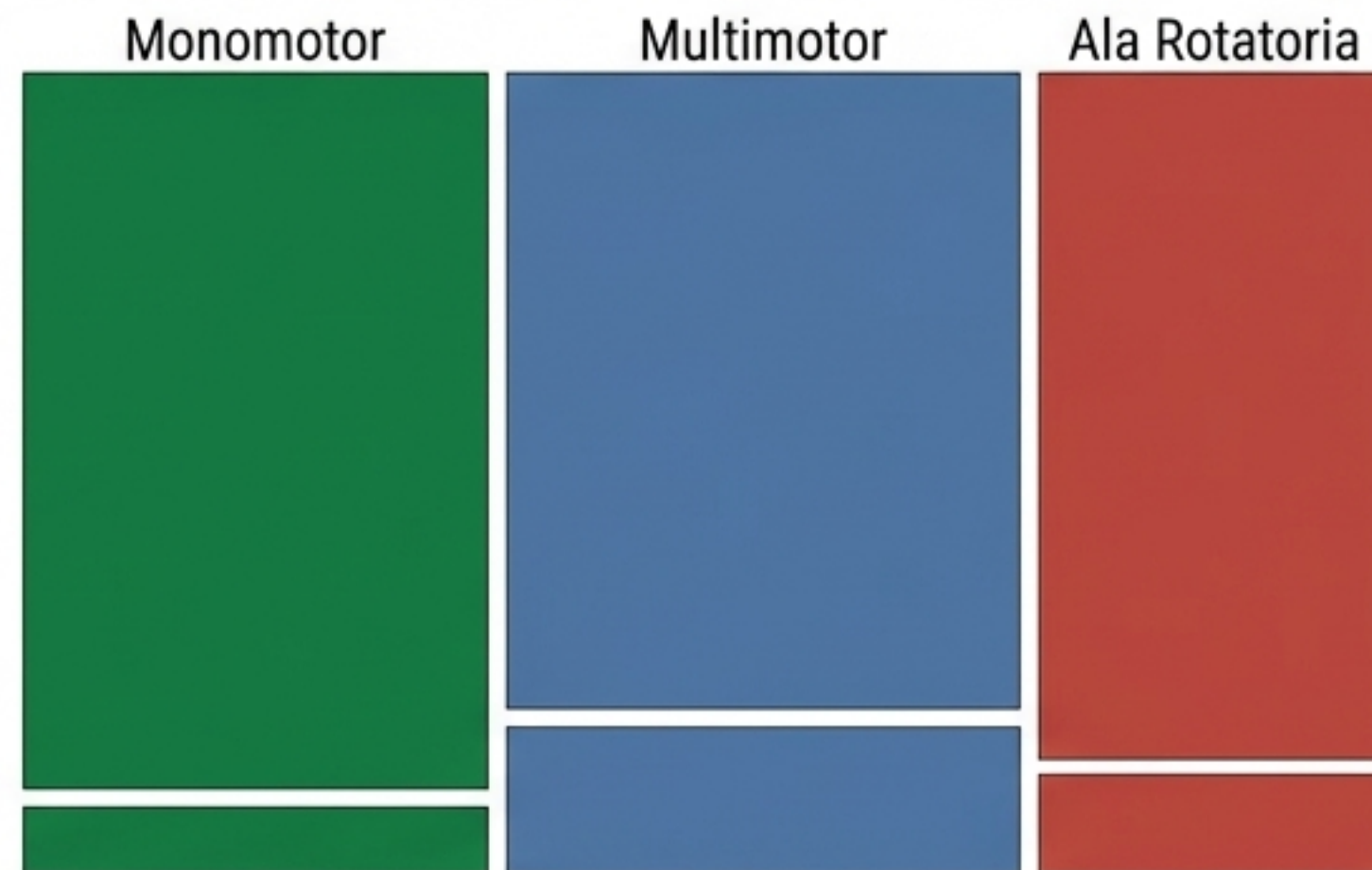


El ancho de la columna representa el volumen total de esa categoría.

WIDTH_FACTOR: TOTAL_VOLUME_REPRESENTATION

Paso 2: División Secundaria

DATA_STREAM: SECONDARY_SPLIT_PROCESS



La altura de las subdivisiones representa la proporción interna.

HEIGHT_FACTOR: INTERNAL_PROPORTION_VISUALIZATION

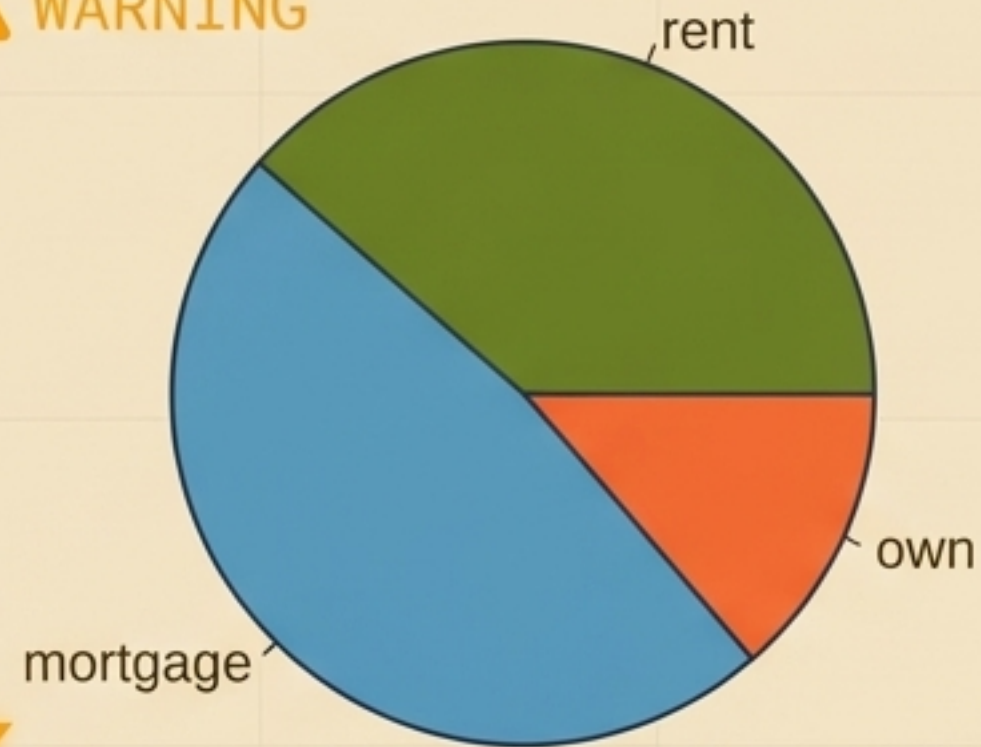
SÍNTESIS ANALÍTICA ⚠

Un gráfico de mosaico es un gráfico apilado estandarizado que recupera la **información perdida del volumen** a través del **área geométrica**. Muestra asociaciones sin ocultar los **tamaños relativos de la flota**.

La Trampa del Gráfico de Pastel: Un Riesgo Cognitivo

EL PROBLEMA

! WARNING



El cerebro humano procesa mal los ángulos y las áreas circulares. Requiere tiempo adicional descifrar si la sección Monomotor es mayor o menor que Multimotor.

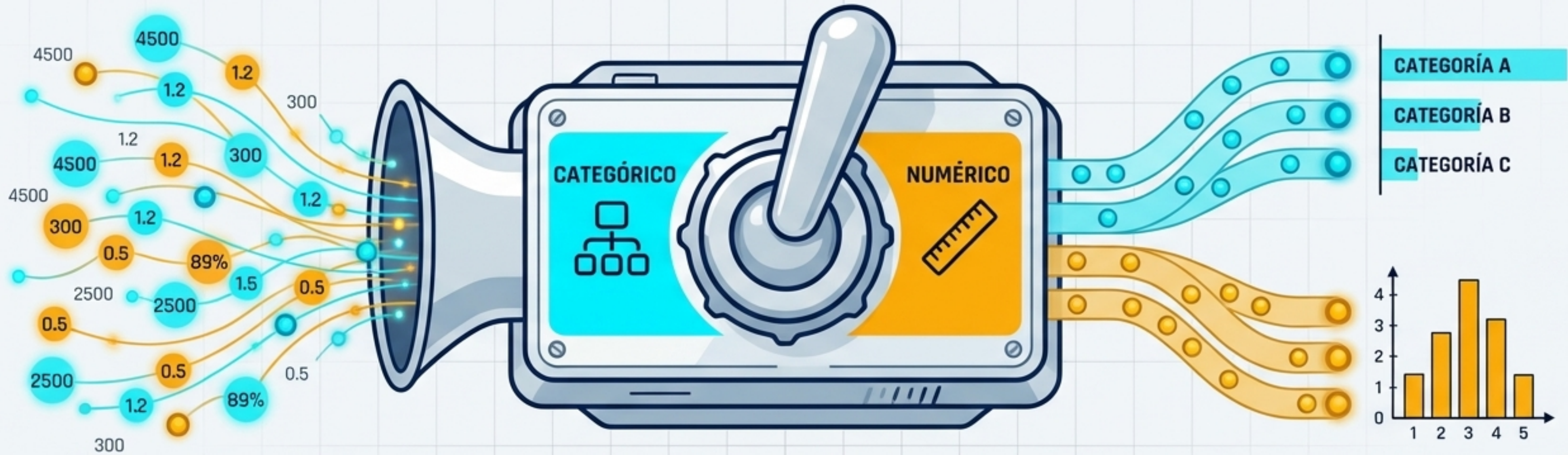
LA SOLUCIÓN



La alineación lineal permite una comparación instantánea e inequívoca. Multimotor es claramente superior.



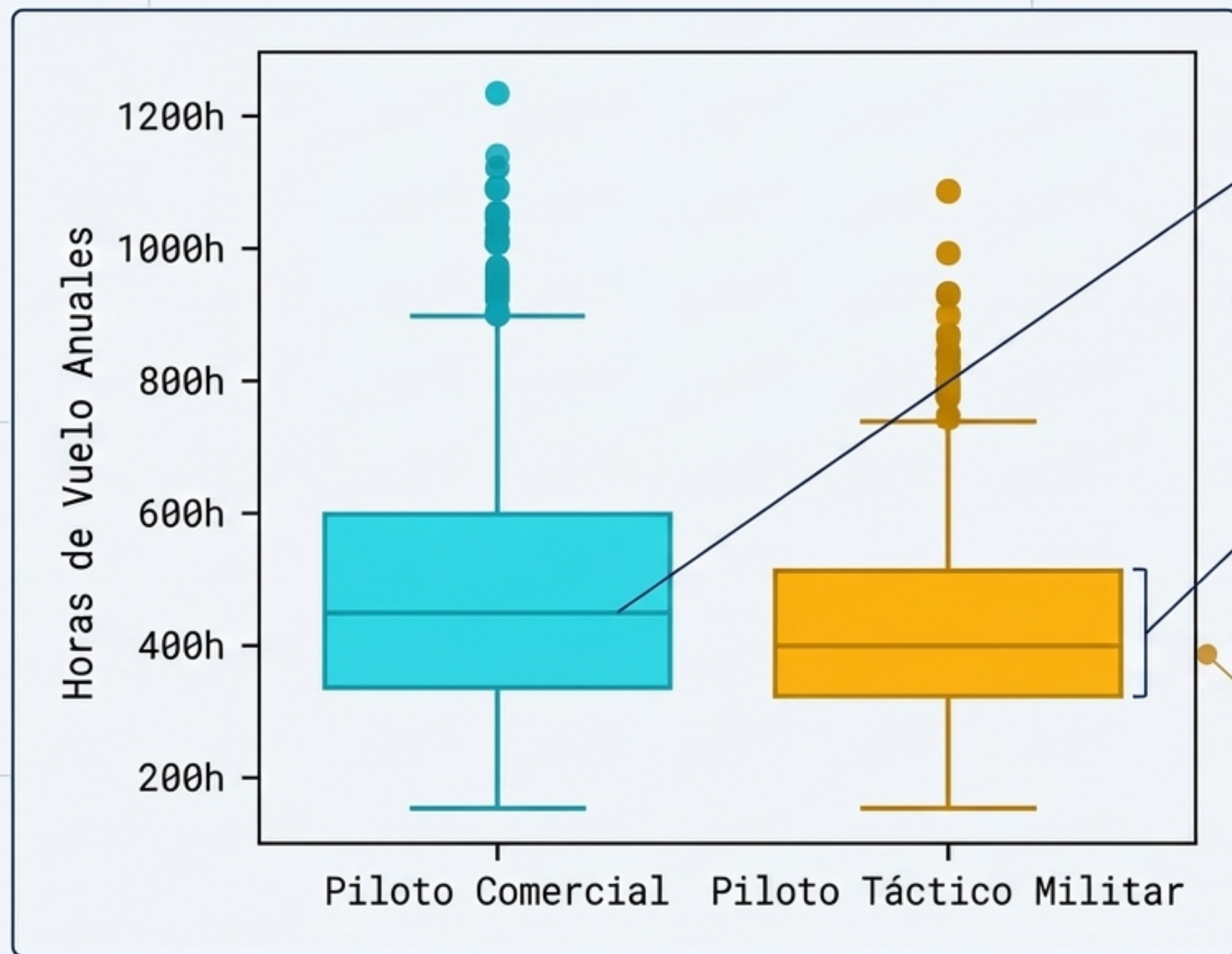
PROTOCOLO APROBADO: Para reportes operacionales críticos, reemplace siempre los gráficos de pastel por gráficos de barras.



Cruzando la Frontera: Categórico vs. Numérico

- Los análisis más profundos ocurren en la intersección de distintos tipos de datos.
- ¿Qué sucede cuando queremos comparar una métrica continua (como las Horas de Vuelo Anuales) a través de una agrupación categórica (como Rango de Piloto)?
- Sistemas Desplegados: Gráficos de Cajas y Histogramas Huecos.

Gráficos de Cajas Lado a Lado: Comparando Distribuciones

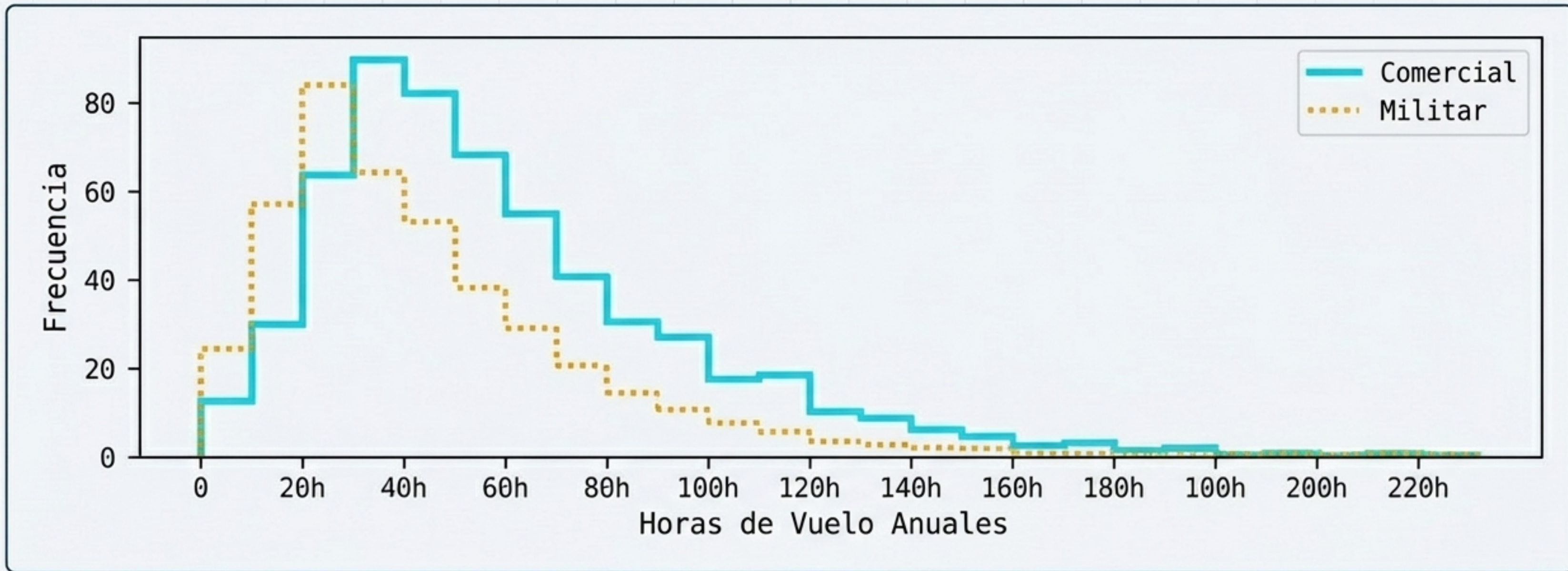


Centros: La mediana comercial (~450 horas) es superior a la militar (~400 horas).

Dispersión (Rango Intercuartílico): La variabilidad es notablemente mayor (un 50% más amplia) en el sector comercial, indicado por la altura de la caja.

Anomalías (Outliers): Ambos grupos presentan pilotos de ultra-largo radio muy por encima de los bigotes superiores.

Histogramas Huecos: Revelando la Forma Oculta

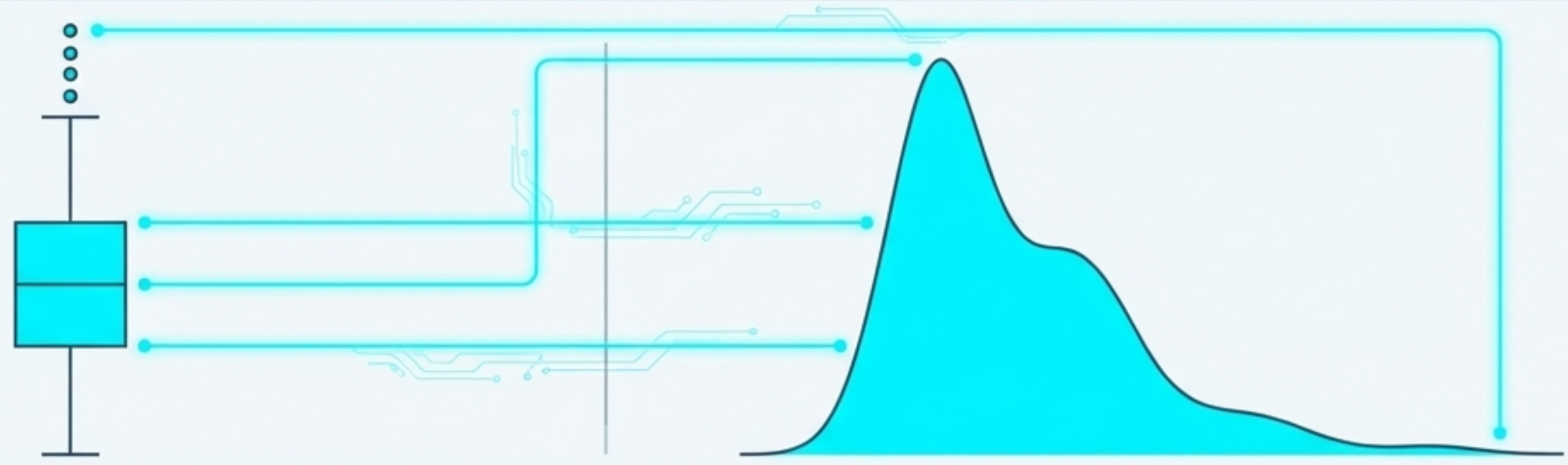


Concepto: Al superponer solo los contornos de los histogramas, podemos comparar grupos sin que los colores sólidos bloqueen la información.

Hallazgos Operativos: A diferencia del gráfico de cajas, aquí observamos claramente la forma de la distribución: ambos grupos presentan una ligera asimetría hacia la derecha (right skew) y son unimodales (un solo pico central dominante).

Gráfico de Cajas (Box Plot)

Histograma



Línea Central -> El Pico

La línea de la mediana en el gráfico de cajas se alinea cerca (pero no exactamente) de la moda principal (el pico) en el histograma.

La Caja (IQR) -> La Masa

El 50% central de los datos se concentra donde el histograma es más denso.

Bigotes y Puntos -> La Cola

Los outliers superiores del gráfico de cajas mapean directamente a la larga cola derecha que se arrastra en el histograma.

El gráfico de cajas es superior para comparar centros y dispersión. El histograma es superior para descubrir formas, sesgos y anomalías ocultas.

Síntesis de Herramientas Analíticas: Matriz Operacional

Una Variable Categórica

- **Herramientas:** Gráfico de Barras Simple, Tabla de Frecuencias/Proporciones.
- **Uso:** Evaluar volumen total o participación de mercado de la flota.

Dos Variables Categóricas

- **Herramientas:** Gráfico de Barras Apilado, Lado a Lado, y Estandarizado.
- **Uso:** Cruzar variables (ej. Tipo de Avión por Background del Piloto) para detectar asociaciones directas.

Datos Multidimensionales Complejos

- **Herramientas:** Gráfico de Mosaico.
- **Uso:** Visualizar proporciones y volumen total simultáneamente sin perder contexto de tamaño.

Categorico Cruzado con Numérico

- **Herramientas:** Gráficos de Cajas Lado a Lado, Histogramas Huecos.
- **Uso:** Comparar métricas de rendimiento continuo (Horas, Ingresos) segregadas por categorías estructurales.