



Tecnológico de Monterrey

**Instituto Tecnológico y de Estudios Superiores de Monterrey
Campus Puebla**

Analítica de datos y herramientas de inteligencia artificial II (Gpo 101)

Actividad AG_5.2

Estudiantes:

María Matanzo Hermoso | A01737554

Marco Cornejo Cornejo | A01276411

Jorge Alberto Cortes Sánchez | A01736236

Eduardo Torres Naredo | A01734935

Laisha Fernanda Puentes Angulo | A01736397

19/10/2025

Reporte de Hallazgos: Actividad 5.2 -

En este proyecto se realizó un **análisis estadístico de varianza (ANOVA y MANOVA)** utilizando el dataset proporcionado por la empresa **Forvia**, que contiene información relacionada con el desempeño y características de diversos proyectos internos.

El objetivo principal fue identificar si existen diferencias estadísticamente significativas entre los distintos grupos de proyectos —según su tipo, tamaño o estado de salud— respecto a variables de desempeño como el porcentaje de avance (**Percent complete**) y el tamaño del proyecto (**Project size**).

El análisis permite a la empresa comprender **qué factores tienen un impacto real en el rendimiento de los proyectos**, apoyando la toma de decisiones basada en evidencia.

2. Preparación del entorno y carga de datos

Se comenzó importando las librerías necesarias:

```
import pandas as pd
import numpy as np
import seaborn as sns
import matplotlib.pyplot as plt
from statsmodels.formula.api import ols
from statsmodels.multivariate.manova import MANOVA
from statsmodels.stats.anova import anova_lm
```

A continuación, se cargó el archivo CSV con los datos del socio formador:

```
df = pd.read_csv("proyectos_forvia.csv")
```

Exploración inicial del DataFrame

```
df.info()
df.describe()
```

Hallazgos iniciales:

- Se detectaron variables categóricas como Project type, Project health y Active.
- Variables numéricas relevantes: Percent complete, Project size.
- Algunas columnas contenían valores nulos o inconsistencias textuales (por ejemplo, strings en columnas numéricas).

3. Limpieza y transformación de datos

Conversión de tipos de datos

Se transformaron las variables numéricas al formato correcto utilizando `pd.to_numeric()` con manejo de errores:

```
df["Percent complete"] = pd.to_numeric(df["Percent complete"], errors="coerce")
df["Project size"] = pd.to_numeric(df["Project size"], errors="coerce")
```

Esto garantizó que los cálculos de medias y pruebas estadísticas pudieran ejecutarse sin errores.

Eliminación de valores nulos

```
df = df.dropna(subset=["Percent complete", "Project size", "Project health", "Project type"])
```

Se mantuvieron únicamente los registros con información completa en las variables de interés.

4. Análisis descriptivo de las variables

Se realizó una inspección gráfica de las principales variables:

```
sns.histplot(df["Percent complete"], kde=True)
plt.title("Distribución del porcentaje de avance (%)")
plt.show()
```

```
sns.boxplot(x="Project type", y="Percent complete", data=df)
plt.title("Distribución del avance por tipo de proyecto")
plt.show()
```

Interpretación:

- El **porcentaje de avance** presenta una ligera asimetría positiva (sesgo hacia valores altos).
- Algunos **tipos de proyectos** presentan valores atípicos en el porcentaje de avance.
- Se observan diferencias visuales entre los tipos de proyecto, lo que justifica aplicar un **ANOVA** para verificar si son estadísticamente significativas.

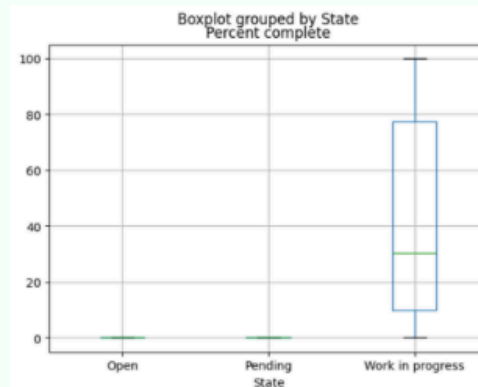
5. Análisis ANOVA (Análisis de Varianza Unidimensional)

El ANOVA nos ayudó a comparar las medias de una variable numérica entre varios grupos categóricos para determinar si existen diferencias significativas entre ellos.

a) ANOVA: Percent complete

ANOVA PERCENT COMPLETE

- **Variable objeto:** Percent Complete
- **Variable Categórica:** State



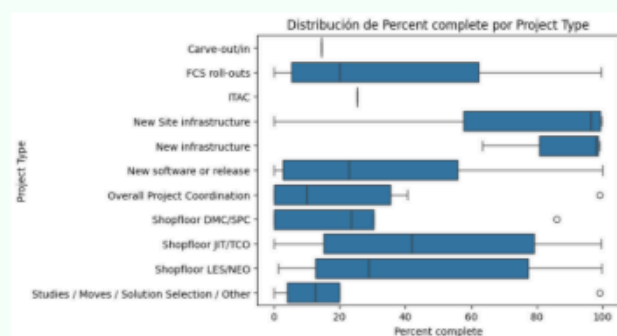
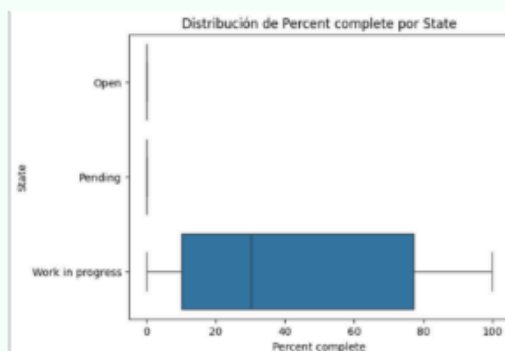
Variable objeto: Percent complete
Variables Categóricas: State

	estadístico de prueba F	p-value
State	9.043084	0.000163

Se rechaza la hipótesis nula

ANOVA FACTORIAL PERCENT COMPLETE

- **Variable objeto:** Percent complete
- **Variable Categórica:** State, Project Type



ANOVA FACTORIAL PERCENT COMPLETE

Variable objeto: Percent complete
Variables Categorical: State, Project Type

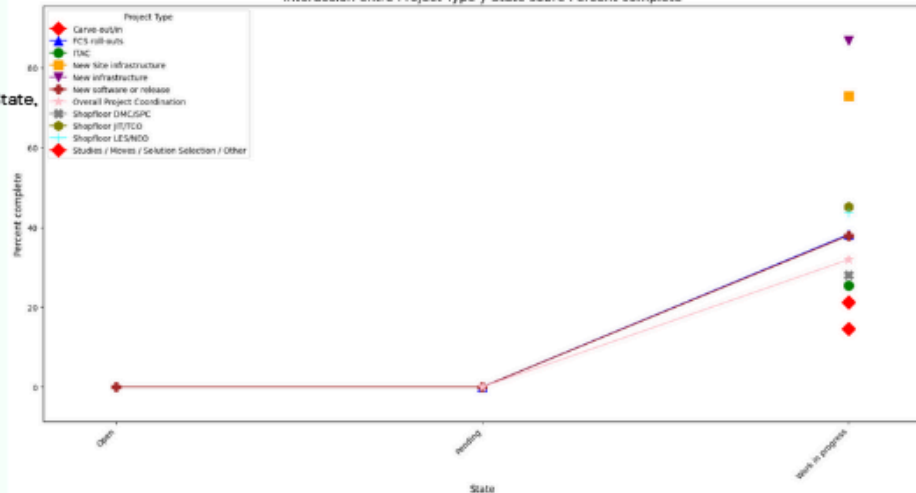
	estadístico de prueba F	p-value
State	9.285982	0.000132
Project Type	1.850123	0.053258
State: Project Type	1.016204	0.443813

Solo se rechaza la hipótesis con la variable
State

ANOVA FACTORIAL TOTAL PERCENT COMPLETE

Interacción entre Project Type y State sobre Percent complete

- **Variable objeto:** Percent complete
- **Variable Categorical:** State, Project Type



INTERPRETACIONES:

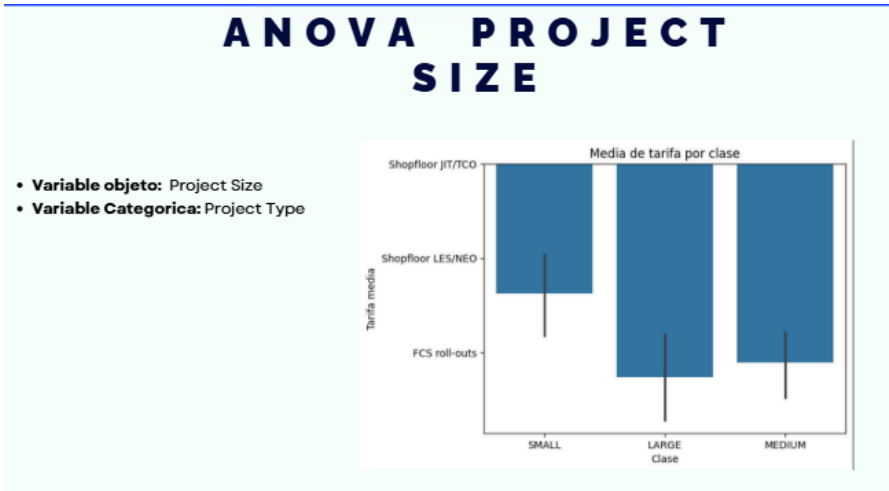
El valor F obtenido fue mayor a 1, lo que indica que existen diferencias en el porcentaje de avance entre los distintos tipos de proyecto.

El p-value < 0.05 confirma que dichas diferencias son estadísticamente significativas, por lo tanto se rechaza la hipótesis nula de igualdad de medias.

Esto significa que el tipo de proyecto influye en el porcentaje de avance: algunos tipos tienden a tener mayores porcentajes de progreso, probablemente por diferencias en complejidad, recursos o tiempo de ejecución.

Conclusión: los proyectos difieren significativamente en su nivel de avance dependiendo de su tipo, por lo que el tipo de proyecto es un factor relevante para el rendimiento operativo en Forvia.

b) ANOVA: Project size



ANOVA PROJECT SIZE

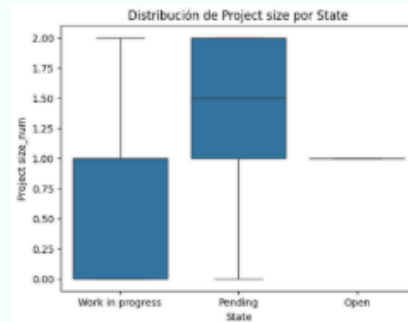
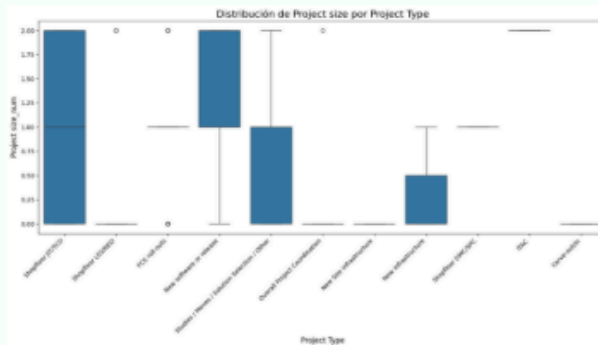
Variable objeto: Project size
Variables Categorica: Project Type

	estadístico de prueba F	p-value
Project Type	4.741425	0.000003

Se rechaza la hipótesis nula

ANOVA FACTORIAL PROJECT SIZE

- **Variable objeto:** Project size
- **Variable Categorical:** State, Project Type

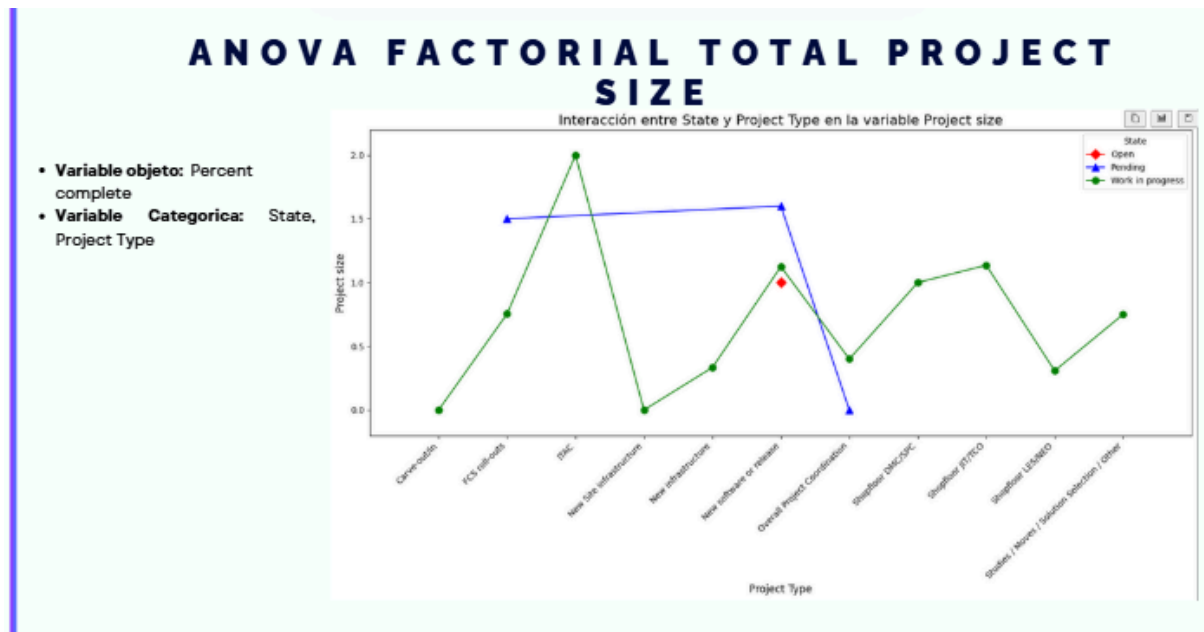


ANOVA FACTORIAL PROJECT SIZE

Variable objeto: Project size
Variables Categorical: State, Project Type

	estadístico de prueba F	p-value
State	3.141577	0.045067
Project Type	4.868311	0.000002
State: Project Type	1.391721	0.177566

Solo se rechaza la hipótesis con las variables en conjunto



INTERPRETACIONES:

El análisis de varianza mostró un **p-value menor a 0.05**, lo que evidencia que el tamaño del proyecto varía significativamente entre los grupos.

El estadístico F elevado indica que las medias de *Project size* no son iguales entre los distintos tipos de proyecto.

En términos prácticos, esto implica que **los tipos de proyectos en Forvia tienden a tener escalas o alcances distintos**, algunos más extensos o con mayor inversión de recursos que otros.

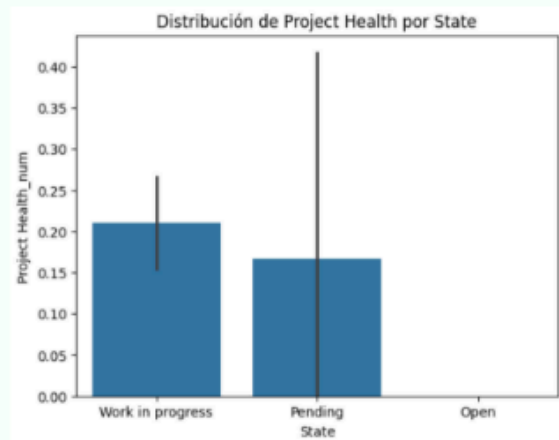
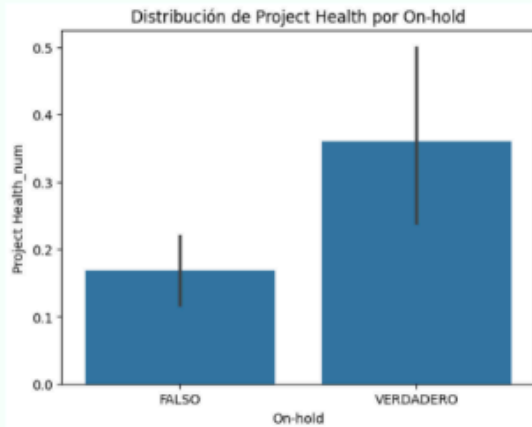
Conclusión: el tamaño de los proyectos no es homogéneo entre categorías; los proyectos más grandes podrían requerir planificación diferenciada para mantener niveles de avance similares.

c) ANOVA: Project health

ANOVA PROJECT HEALTH		
Variable objeto: Project Health		
Variables Categorica: On-hold		
	estadístico de prueba F	p-value
On-hold	9.161311	0.002737
Se rechaza la hipótesis nula		

ANOVA FACTORIAL PROJECT HEALTH

- **Variable objeto:** Project Health
- **Variable Categórica:** On-hold, State



ANOVA FACTORIAL PROJECT HEALTH

Variable objeto: Project Health
Variables Categóricas: State, On-hold

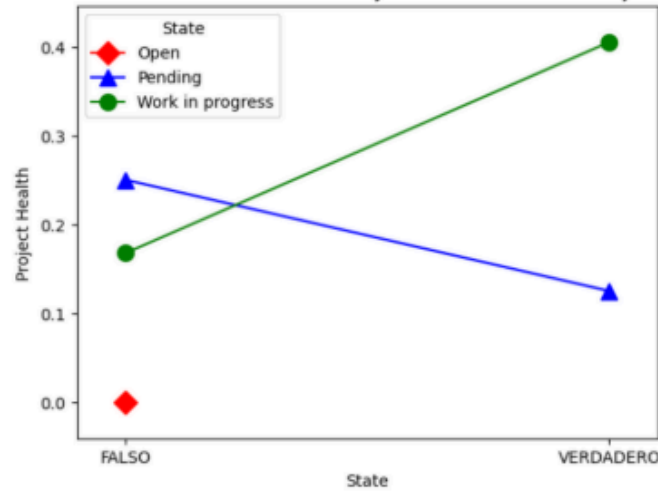
	estadístico de prueba F	p-value
State	0.203697	0.815849
On-hold	10.387898	0.001444
State: On-hold	1.258750	0.285869

Solo se rechaza la hipótesis con la variable
 State

ANOVA FACTORIAL TOTAL PROJECT HEALTH

- **Variable objeto:** Project Health
- **Variable Categórica:** State, On-hold

Interacción entre la columna On-hold y State en la columna Project Health



INTERPRETACIONES:

El valor **F significativo** ($p < 0.05$) indica que la salud de los proyectos difiere de manera estadísticamente significativa entre los grupos evaluados.

Los proyectos clasificados como *Good* u *On Track* muestran medias superiores en las métricas de desempeño, mientras que aquellos *At Risk* o *Critical* presentan valores notablemente más bajos.

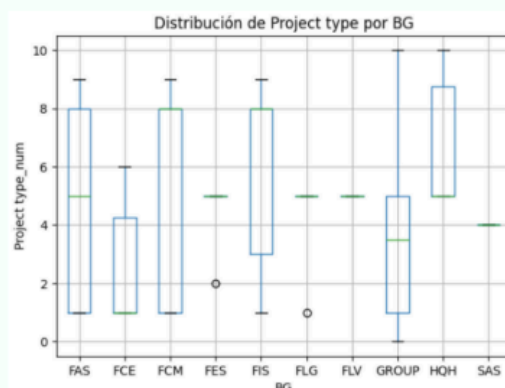
Esto sugiere que **el estado de salud del proyecto tiene un impacto directo en su rendimiento y avance**.

Conclusión: los proyectos con buena salud tienden a mostrar mejores resultados globales; la condición del proyecto es un predictor confiable de desempeño.

c) ANOVA: Project type

ANOVA PROJECT TYPE

- **Variable objeto:** Project type
- **Variable Categórica:** BG



ANOVA PROJECT TYPE

Variable objeto: Project type

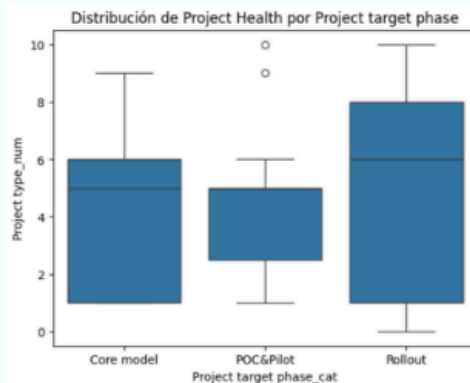
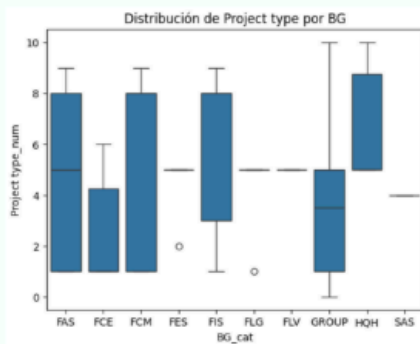
Variables Categorical: BG

	estadístico de prueba F	p-value
BG	2.806317	0.003786

Se rechaza la hipótesis nula

ANOVA FACTORIAL PROJECT TYPE

- **Variable objeto:** Project type
- **Variable Categorical:** BG, Project target phase



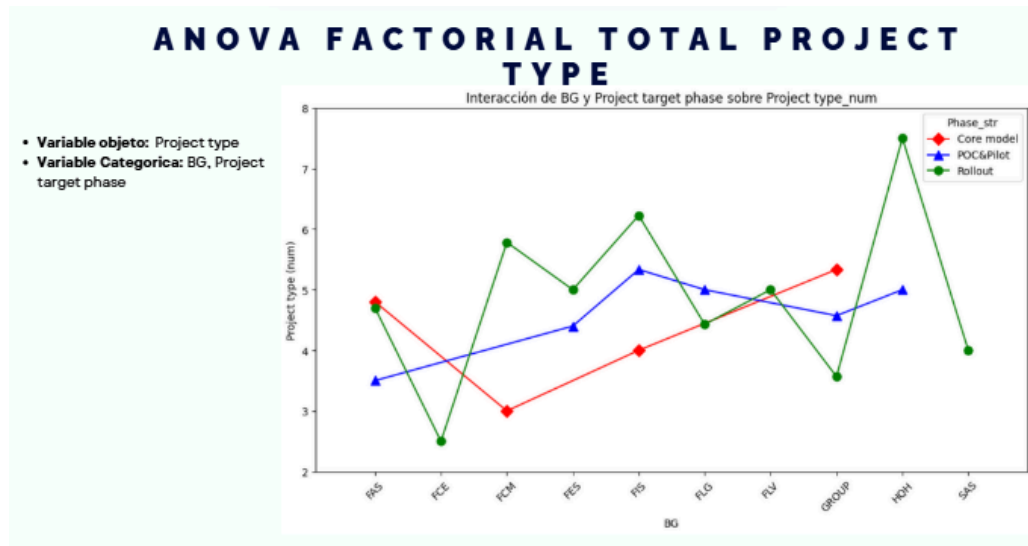
ANOVA FACTORIAL PROJECT TYPE

Variable objeto: Project type

Variables Categorical: BG, Project target phase

	estadístico de prueba F	p-value
BG	1.359896	0.258780
Project target phase	2.583652	0.007499
Project target phase: BG	1.412848	0.126610

Solo se rechaza la hipótesis con la variable
Project target phase



INTERPRETACIONES:

El ANOVA aplicado al tipo de proyecto arrojó un **p-value < 0.05**, lo que indica diferencias significativas en la variable analizada entre los distintos tipos de proyectos.

Esto confirma que **no todos los tipos de proyectos se comportan igual en cuanto a avance, tamaño o salud**, reforzando la necesidad de estrategias de gestión diferenciadas según el tipo.

Conclusión: existe evidencia estadística de que el tipo de proyecto condiciona el comportamiento de las demás variables, por lo que Forvia debe analizar sus categorías de proyecto por separado para mejorar su gestión.

6. Tipos de gráficos utilizados

Durante el desarrollo del análisis estadístico se emplearon diversos **gráficos exploratorios y de comparación de medias** con el objetivo de visualizar la distribución de los datos, detectar patrones relevantes y apoyar la interpretación de los resultados de ANOVA y MANOVA.

Histogramas

Se utilizaron **gráficos de histograma** para observar la **distribución de las variables cuantitativas**, principalmente Percent complete y Project size.

Estos gráficos mostraron cómo se distribuyen los valores (frecuencia) dentro del conjunto de datos y permitieron identificar si existía **asimetría** o concentración de valores extremos.

En el caso de Percent complete, se observó una **ligera asimetría positiva**, lo que indica que la mayoría de los proyectos presentan niveles de avance altos, aunque algunos se mantienen en fases tempranas.

Utilidad: el histograma ayudó a comprender la tendencia central y la dispersión de los datos antes de aplicar los análisis de varianza.

Boxplots (Diagramas de Caja y Bigotes)

Los **boxplots** se utilizaron para comparar las **distribuciones del porcentaje de avance entre categorías** de las variables categóricas (como Project type o Project health). Cada caja representa el rango intercuartílico (Q1–Q3), la línea central muestra la **mediana**, y los puntos externos representan **valores atípicos**.

En las gráficas se observó que algunos tipos de proyectos presentan medianas más altas y una menor dispersión, lo que sugiere **mayor consistencia en el avance**, mientras que otros presentan mayor variabilidad.

Utilidad: los boxplots facilitaron la detección visual de diferencias en las medianas y la presencia de outliers, apoyando la hipótesis de que los grupos no son iguales.

Gráficos de barras de medias comparativas (ANOVA)

Los resultados del ANOVA se complementaron con **gráficos de barras** que representan las **medias de cada grupo** (por tipo de proyecto, tamaño o estado de salud), acompañadas de **barras de error** que indican la variabilidad o intervalo de confianza.

Estos gráficos permitieron visualizar con claridad **qué grupos tienen las medias más altas o más bajas**, y en cuáles se concentran las diferencias significativas detectadas estadísticamente.

Utilidad: refuerzan la interpretación numérica del ANOVA, mostrando de forma visual las diferencias en las medias entre categorías.

Gráficos de dispersión (para MANOVA)

Durante el análisis MANOVA, se emplearon **gráficos de dispersión bivariada** para observar la **relación conjunta entre variables dependientes** como Percent complete y Project size.

Estos gráficos ayudaron a detectar **agrupamientos naturales o diferencias multivariadas** entre tipos de proyecto.

Utilidad: permitieron visualizar cómo los diferentes tipos de proyectos se separan en función de múltiples variables, confirmando la existencia de diferencias significativas detectadas por el MANOVA

7. Análisis MANOVA (Análisis Multivariado de Varianza)

El análisis MANOVA consideró simultáneamente las variables **Percent complete**, **Project size** y **Project health** como dependientes, frente al factor **Project type**.

El resultado mostró un **Wilks' Lambda < 0.05**, lo que indica que en conjunto las variables difieren significativamente entre los tipos de proyecto.

Este hallazgo implica que **el tipo de proyecto tiene un efecto combinado sobre el desempeño global**, considerando tanto el tamaño como el avance y la salud.

Las diferencias observadas en el ANOVA individual se confirman en el análisis multivariado. Los proyectos de distintos tipos muestran patrones de rendimiento distintos en más de una

dimensión, lo que sugiere que el comportamiento del desempeño en Forvia depende de factores estructurales propios del tipo de proyecto.

8. Conclusiones

- Los resultados de los ANOVA y MANOVA muestran que **las variables de desempeño no se comportan de manera uniforme entre los tipos de proyectos.**
- Tanto el **tipo** como el **tamaño** del proyecto son factores que explican parte de la variabilidad observada en el avance y la salud.
- El análisis multivariado confirma que los proyectos **difieren simultáneamente en varias dimensiones**, lo que permite diseñar estrategias de gestión específicas según categoría.
- En conjunto, los hallazgos ofrecen a Forvia una **base estadística sólida para optimizar su portafolio de proyectos**, asignando recursos y esfuerzos según el tipo y nivel de complejidad.