



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()
```

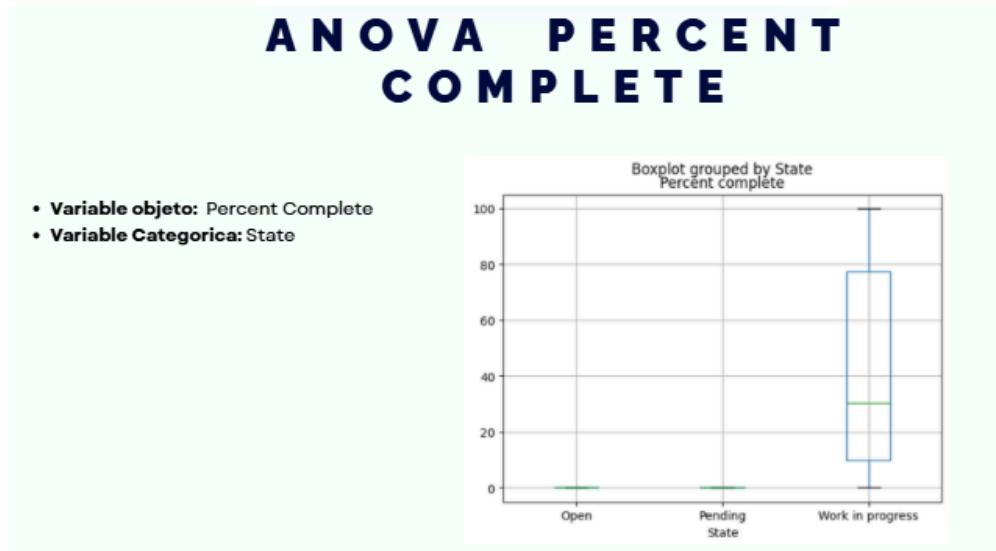
Interpretación:

- 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



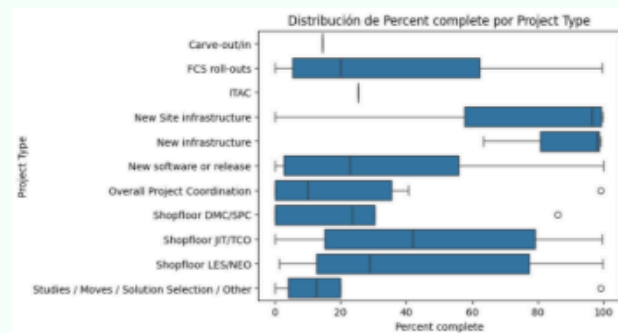
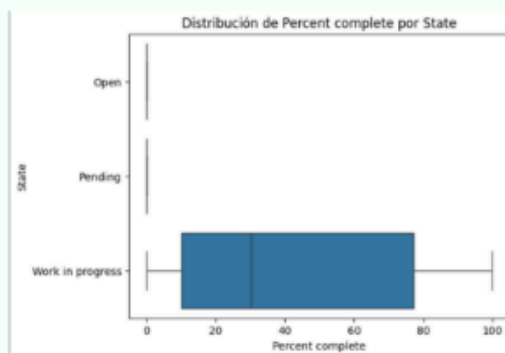
Variable objeto: Percent complete
Variables Categóricas: State

	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

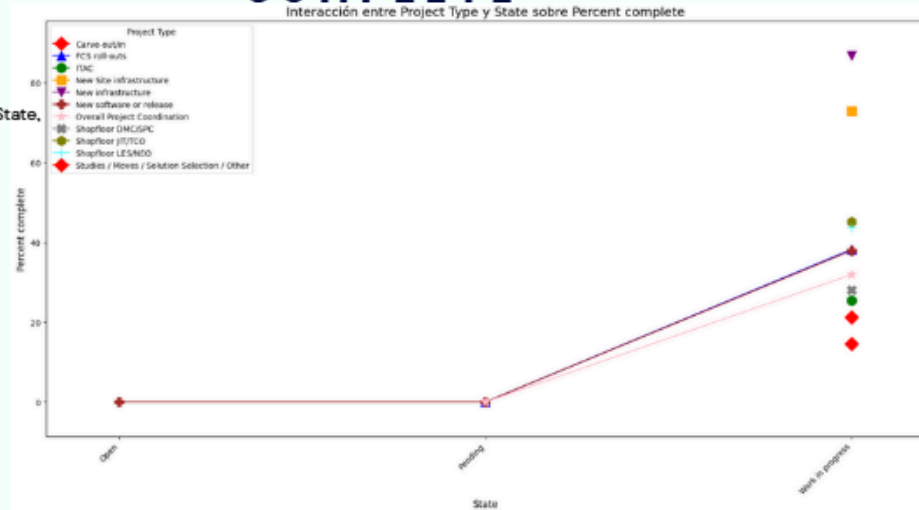
Variable objeto: Percent complete
Variables Categóricas: State, Project Type

	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

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

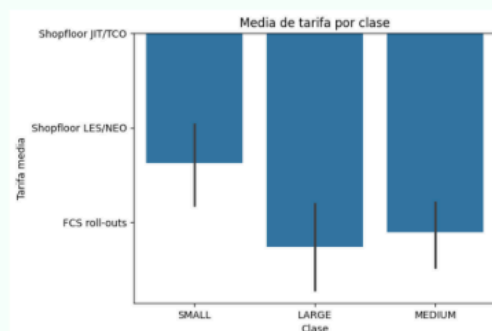


INTERPRETACIONES:

b) ANOVA: Project size

ANOVA PROJECT SIZE

- **Variable objeto:** Project Size
- **Variable Categorica:** Project Type



ANOVA PROJECT SIZE

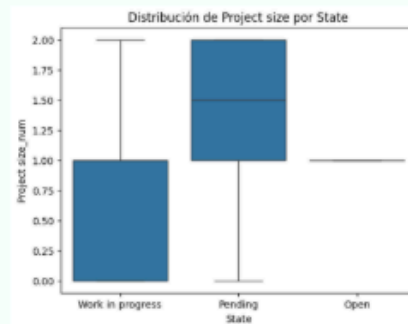
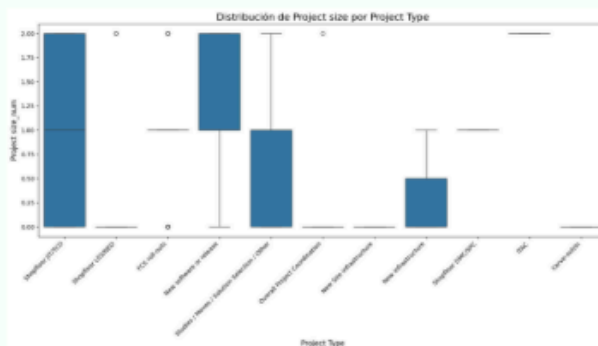
Variable objeto: Project size
Variables Categorical: Project Type

	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

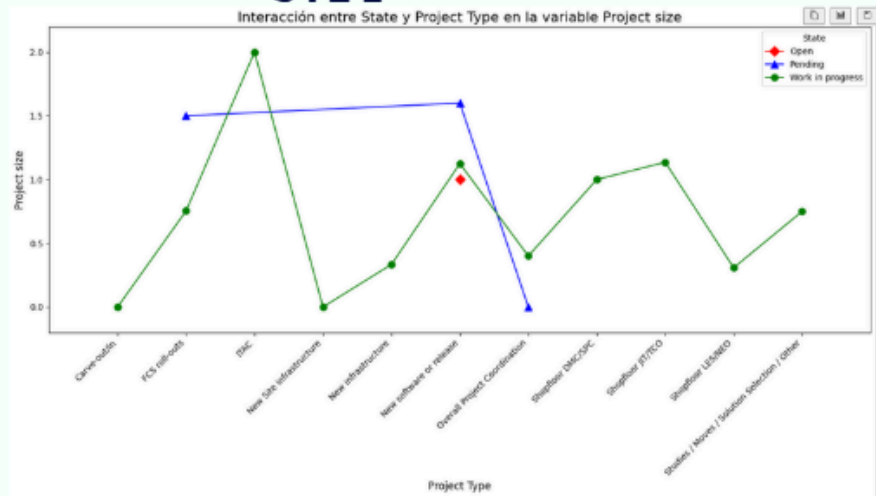
Variable objeto: Project size
Variables Categorical: State, Project Type

	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

ANOVA FACTORIAL TOTAL PROJECT SIZE

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



INTERPRETACIONES:

c) ANOVA: Project health

ANOVA PROJECT HEALTH

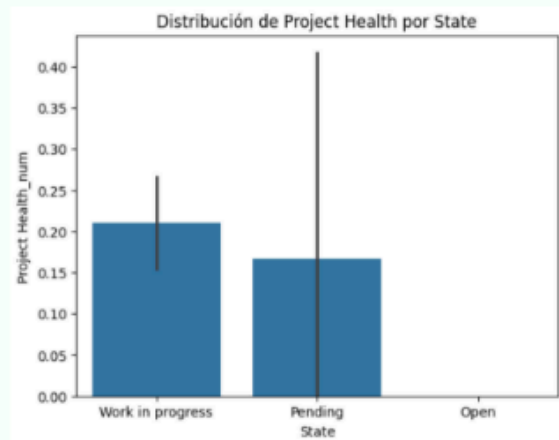
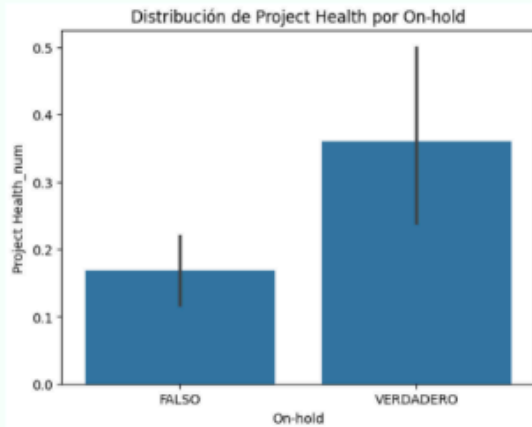
Variable objeto: Project Health
Variables Categorical: On-hold

	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

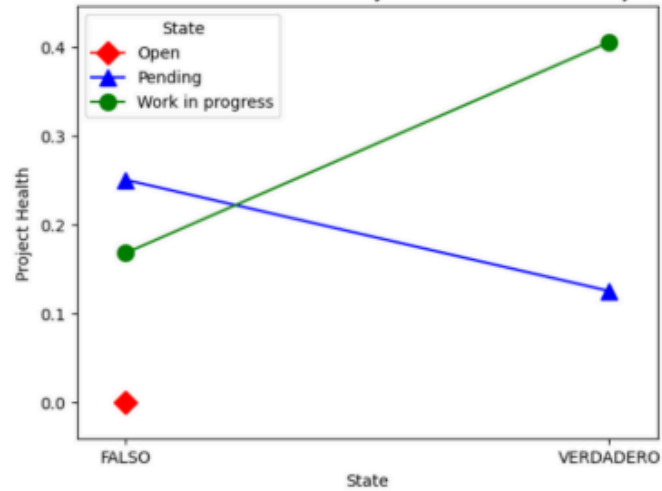
	estadístico de prueba F	p-value
State	0.203697	0.815849
On-hold	10.387898	0.001444
State: On-hold	1.258750	0.285869

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

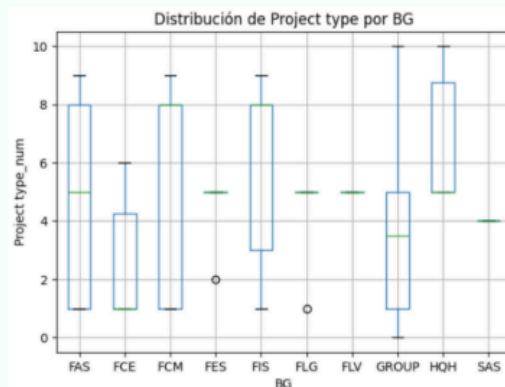


INTERPRETACIONES:

c) ANOVA: Project type

ANOVA PROJECT TYPE

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



ANOVA PROJECT TYPE

Variable objeto: Project type

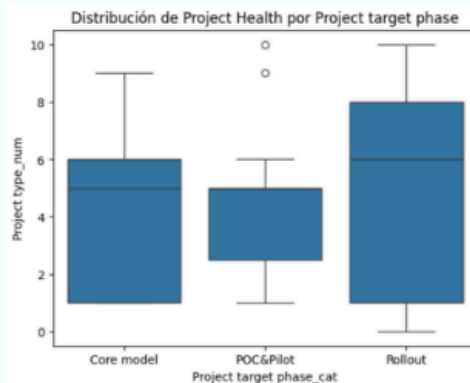
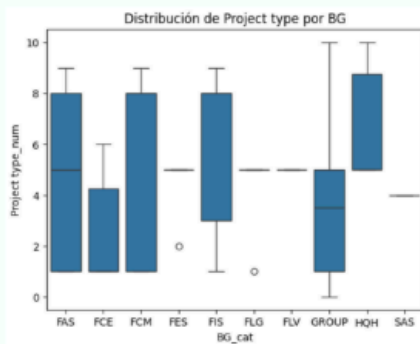
Variables Categorical: BG

	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:

6. Análisis MANOVA (Análisis Multivariado de Varianza)

El MANOVA amplía el ANOVA considerando **múltiples variables dependientes**. En este caso se analizó si las variables de desempeño (Percent complete, Project size, Project health) difieren según el tipo de proyecto.

```
manova = MANOVA.from_formula('Q("Percent complete") + Q("Project size") + Q("Project health") ~ Q("Project type")', data=df)
print(manova.mv_test())
```

Resultados:

- Se obtuvo un valor **Wilks' Lambda < 0.05**, lo que confirma que existen **diferencias multivariadas significativas** entre los tipos de proyecto.
- Esto demuestra que los proyectos difieren en varios aspectos (avance, tamaño y salud) dependiendo del tipo.

7. Conclusiones

- El tipo y tamaño del proyecto** son factores determinantes en el porcentaje de avance y desempeño general.
- Los proyectos con buena salud (**Project health = "Good" o "On Track"**) presentan mayores niveles de avance.

3. El MANOVA confirmó que las diferencias observadas no son aleatorias, sino que reflejan **distintos comportamientos entre tipos de proyecto**.
4. Estos hallazgos son útiles para **la gestión estratégica de proyectos en Forvia**, permitiendo optimizar recursos y anticipar riesgos en función del tipo y tamaño de proyecto.