***Proyecto final Análisis de datos***

1.- PREGUNTAR

¿Qué tema estás explorando?

Efectividad de los proyectos extracurriculares en relación a las horas docentes pagadas.

¿Cuál es el problema que intentas resolver?

¿Cómo están divididas las horas temporales y de fortalecimiento en cada plantel y su impacto en los indicadores escolares?

¿Qué métricas utilizarás para medir tus datos a fin de lograr tu objetivo? ¿Quiénes son los interesados?

Porcentaje de horas de asesorías vs porcentaje de reprobación. Dirección general, administrativa y académica.

¿Cuál es tu público para este análisis y cómo esto afecta tu proceso de análisis y la presentación?

Directores de área y de planteles. El análisis debe ser sencillo de seguir y la presentación visual y corta para entender las diferencias entre planteles.

¿De qué manera estos datos ayudarán a los interesados a tomar decisiones?

A decidir cuales proyectos tienen prioridad para ser autorizados y cuales tienen poco o no impacto en los indicadores.

2.- PREPARAR

¿Dónde se ubican tus datos?

Formulario llenado por los coordinadores con colores y anotaciones por parte de la D.A.

¿Cómo están organizados los datos?

No hay mucha organización, cada coordinador capturó los datos a su propio entendimiento

¿Hay problemas con el sesgo o la credibilidad de estos datos? ¿Tus datos son confiables, originales, integrales, actuales y citados (ROCCC)?

No hay problemas de sesgo o credibilidad. Todos los datos son primarios, el colegio diseña su propia base de datos.

¿Cómo estás abordando la autorización, la privacidad, la seguridad y la accesibilidad?

El manejo de los datos individuales lo tengo sólo yo, a los directores se les presenta los resúmenes de cada plantel en general.

¿Cómo verificaste la integridad de los datos?

Al comparar con las plantillas cargadas en el sistema de RRHH

¿De qué manera te ayuda a responder tu pregunta?

En los datos viene la información requerida para las estadísticas mencionadas en el objetivo.

¿Existe algún problema con los datos?

No están agrupados o actualizados. Viene información en blanco. Se elaboró la base de datos de nuevo desde cero utilizando la información cargada en el sistema de plantillas.

1.- Completar datos de horas, categoría, subcategoría y supervisores.

3.- PROCESAR

¿Qué herramientas eliges y por qué?

Excel y SQL, Excel porque su interface es muy amigable, SQL para practicar BigQuery

¿Has asegurado la integridad de tus datos?

Sí, es mi misma base de datos

¿Qué pasos seguiste para garantizar que tus datos están limpios?

Validación de datos

¿Cómo puedes verificar que tus datos están limpios y listos para analizar?

No hay acentos, ñ, valores en blanco, errores ortográficos; además están agrupados.

¿Documentaste tu proceso de limpieza para poder revisar y compartir estos resultados?

Sí:

**Limpieza de datos con Excel**

Revisar ortografía

Quitar acentos y Ñ

**Limpieza de datos con SQL**

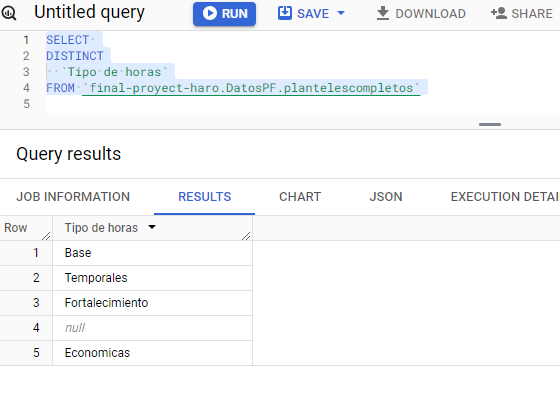
1.- Revisar que no haya errores en tipos de horas

SELECT

DISTINCT

  `Tipo de horas`

FROM `final-proyect-haro.DatosPF.plantelescompletos`



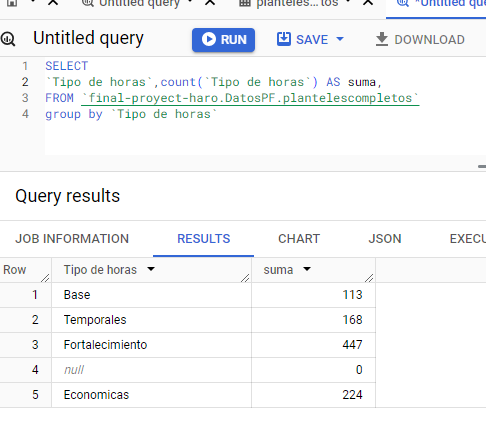
2.- Revisar que no haya valores vacíos

SELECT

`Tipo de horas`,count(`Tipo de horas`) AS suma,

FROM `final-proyect-haro.DatosPF.plantelescompletos`

group by `Tipo de horas`



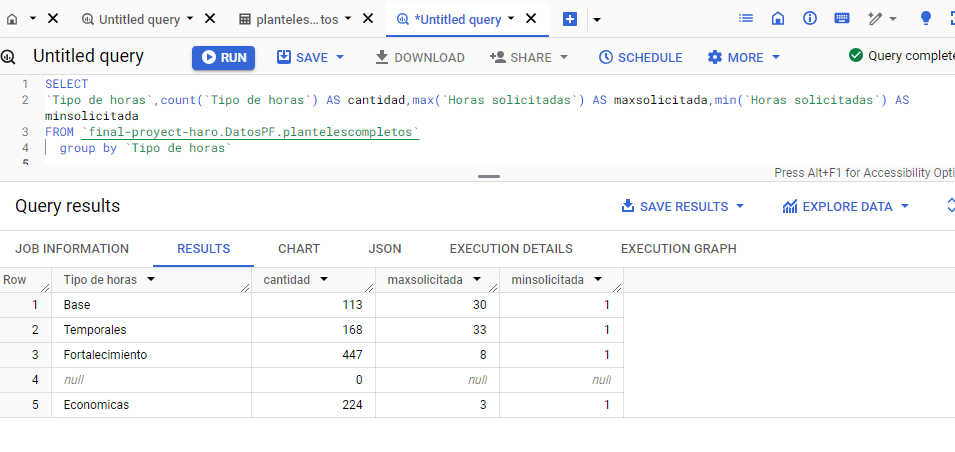
3.- Revisar que no se haya puesto dígitos extra por error, las horas deben ser máximo 8 (excepto casos especiales) y mínimo 1. Fortalecimiento debe ser max 8

SELECT

`Tipo de horas`,count(`Tipo de horas`) AS cantidad,max(`Horas solicitadas`) AS maxsolicitada,min(`Horas solicitadas`) AS minsolicitada

FROM `final-proyect-haro.DatosPF.plantelescompletos`

  group by `Tipo de horas`



4.- Cambiar a mayúsculas o minúsculas:

  SELECT

  `Plantel`,

  upper(`Apellido paterno`) AS lastname1,

  upper(`Apellido materno`) AS lastname2,

  upper(`Nombre`) AS Name,

  `Categoria`,

  `Subcategoria`,

  `Comentario`,

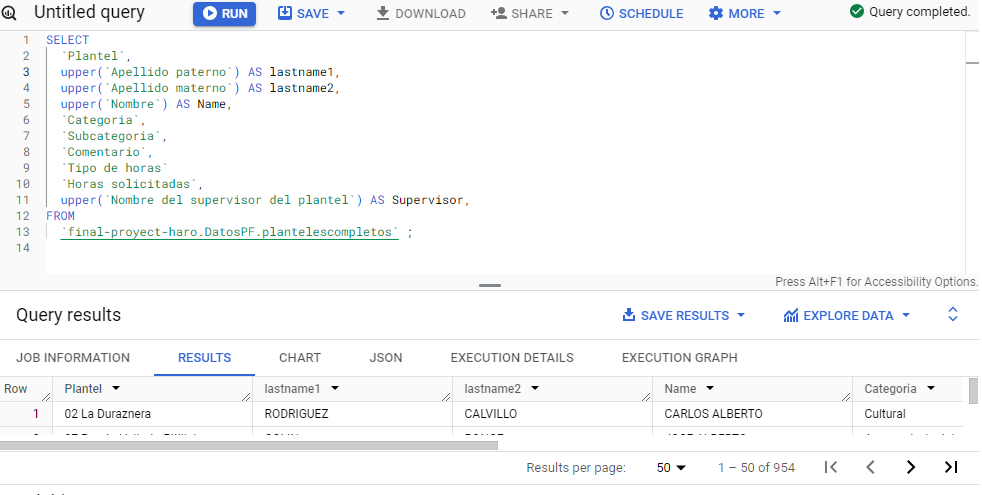
  `Tipo de horas`,

  `Horas solicitadas`,

  upper(`Nombre del supervisor del plantel`) AS Supervisor,

FROM

  `final-proyect-haro.DatosPF.plantelescompletos` ;



5.- Revisar que no haya espacios en blanco

SELECT

  TRIM(`lastname1`) AS Apellidopat,

  TRIM(`lastname2`) AS Apellidomat,

  TRIM(`Name`) AS Nombre,

  `Plantel`,

  TRIM(`Categoria`) AS Category,

  `Subcategoria`,

  TRIM(`Comentario`) AS Comment,

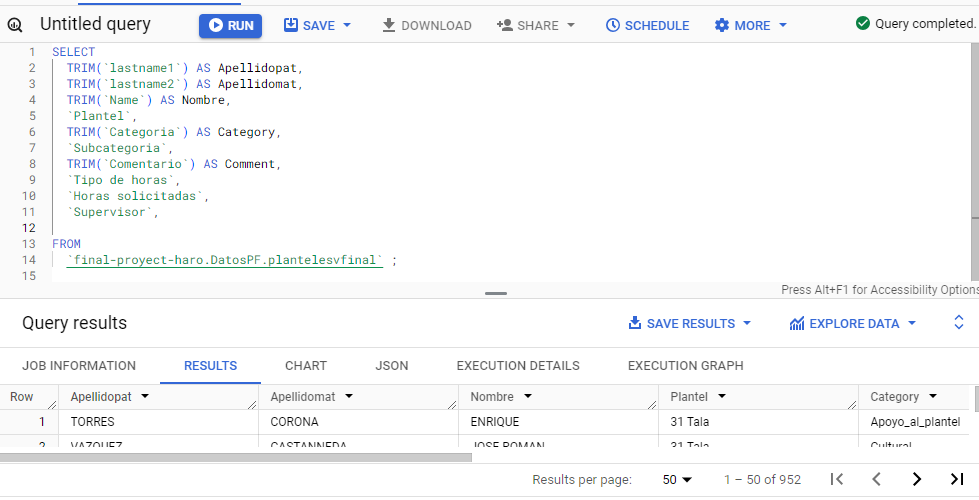
  `Tipo de horas`,

  `Horas solicitadas`,

  `Supervisor`,

FROM

  `final-proyect-haro.DatosPF.plantelesvfinal` ;



4.- ANALIZAR

¿Cómo deberías organizar tus datos para realizar un análisis sobre estos?

Separados en 10 categorías y subcategorías para cada 1 de ellas. Cantidad de horas y tipo (Base, temporal, fortalecimiento o económica)

¿Tus datos se formatearon adecuadamente?

Sí, en la segunda base.

¿Qué sorpresas descubriste en los datos?

Pocas horas dedicadas a asesorías; diferentes cantidades de horas asignadas a la misma actividad; errores de planteles sobre asignación de suficientes tutores escolares y grupales; docentes con cantidades altas de horas, pero sin seguimiento en sus proyectos;

¿Qué tendencias o relaciones encontraste en los datos?

Los planteles foráneos tienen menores índices de reprobación.

¿De qué manera esas conclusiones responden tu pregunta o resuelven tu problema?

Se pudo determinar el porcentaje de asesorías de las horas totales y compararlo con el porcentaje de reprobación.

**Análisis estadístico en R**

install.packages("rmarkdown")

install.packages("tidyverse")

install.packages("janitor")

install.packages("skimr")

install.packages('ggplot2')

library(ggplot2)

library(rmarkdown)

library(tidyverse)

library(janitor)

library(skimr)

datos <- read\_csv("ProyectosR.csv")

colnames(datos)

glimpse(datos)

head(datos)

datos %>% group\_by(Tipo) %>% summarize(mayor=max(Horas), menor=min(Horas),prom=mean(Horas),total=sum(Horas))

tutor<- datos %>% group\_by(Plantel) %>% filter(Subcategoria=="Tutor grupal") %>% summarize(total=sum(Horas))

View(tutor)

ggplot(data = datos, aes(x = Categoria, y = Horas, color = Tipo)) + geom\_point()

ggplot(data = datos, aes(x = Categoria, y = Horas, color = Tipo)) + geom\_point() + facet\_wrap(~Plantel)

ggplot(data = datos, aes(x = Categoria)) + geom\_bar() + facet\_wrap(~Plantel) + theme(axis.text.x = element\_text(angle = 90))

ggplot(data = datos, aes(x = Categoria)) + geom\_bar()+ theme(axis.text.x = element\_text(angle = 90))

ggplot(data = datos, aes(x = Categoria, color=Tipo)) + geom\_bar()+ theme(axis.text.x = element\_text(angle = 90))

apoyo <- datos %>%

filter(Categoria=="Apoyo\_al\_plantel")

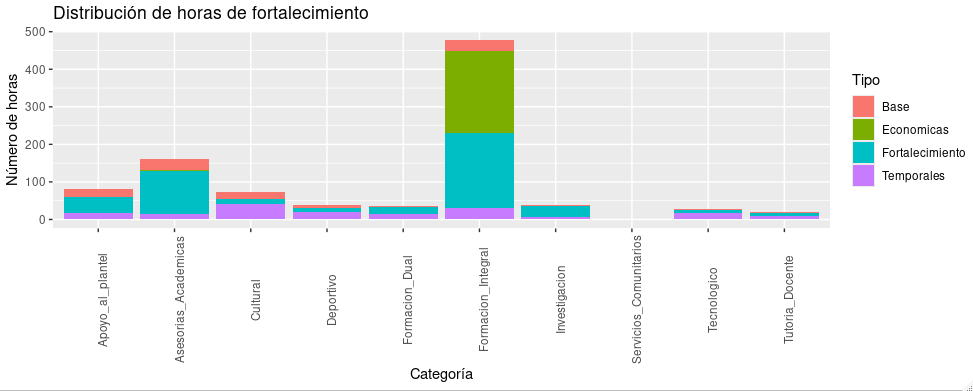
ggplot(data = apoyo, aes(x = Subcategoria, color=Tipo)) + geom\_bar()

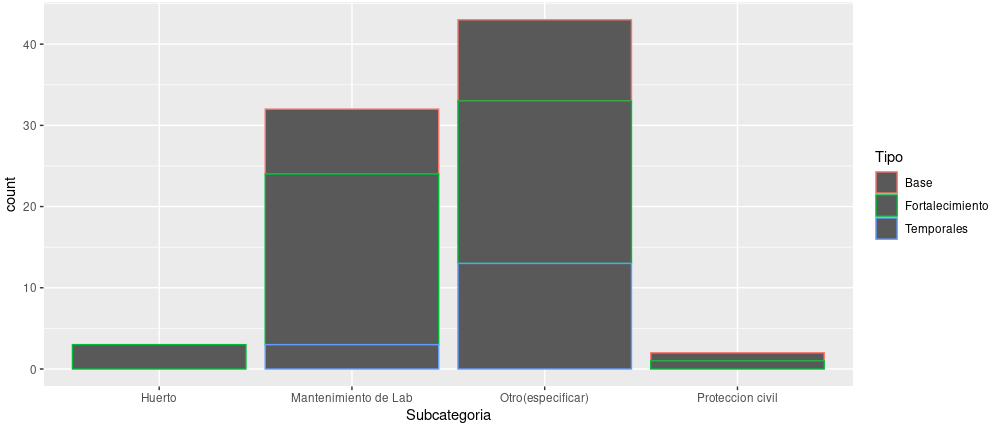
**Gráficas por mostrar**

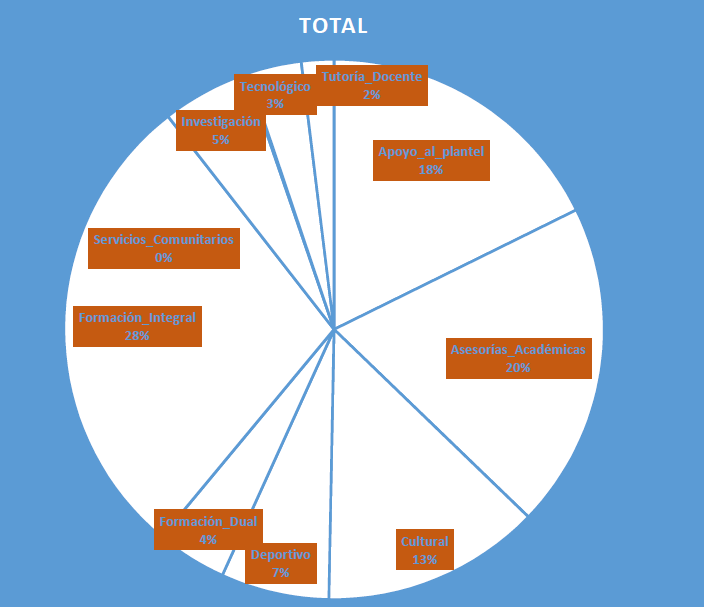
ggplot(data = datos) + geom\_bar(mapping = aes(x = Categoria, fill=Tipo)) + theme(axis.text.x = element\_text(angle = 90))+ labs(title="Distribución de horas de fortalecimiento", x="Categoría", y="Número de horas")

ggplot(data = datos) + geom\_bar(mapping = aes(x = Categoria, fill=Tipo)) + theme(axis.text.x = element\_text(angle = 90))+facet\_wrap(~Plantel)

ggplot(data = apoyo, aes(x = Comentario, color=Tipo)) + geom\_bar() + theme(axis.text.x = element\_text(angle = 90))







porplantela <- datos %>%

filter(Plantel=="01 Tesistan" | Plantel=="02 La Duraznera"| Plantel=="03 Tepatitlan"| Plantel=="04 Cocula"| Plantel=="05 El salto"| Plantel=="06 Totatiche")

ggplot(data = porplantela) + geom\_bar(mapping = aes(x = Categoria, fill=Tipo)) + theme(axis.text.x = element\_text(angle = 90))+ labs(title="Distribución de horas de fortalecimiento", x="Categoría", y="Número de horas") +facet\_wrap(~Plantel)

5.- COMPARTIR

¿Qué historia cuentan tus datos?

La mitad de las horas de fortalecimiento se invierten en tutorías escolares, grupales y orientación educativa

¿De qué manera tus hallazgos se relacionan con tu pregunta original?

Solo el 20% de las horas se invierte en asesorías

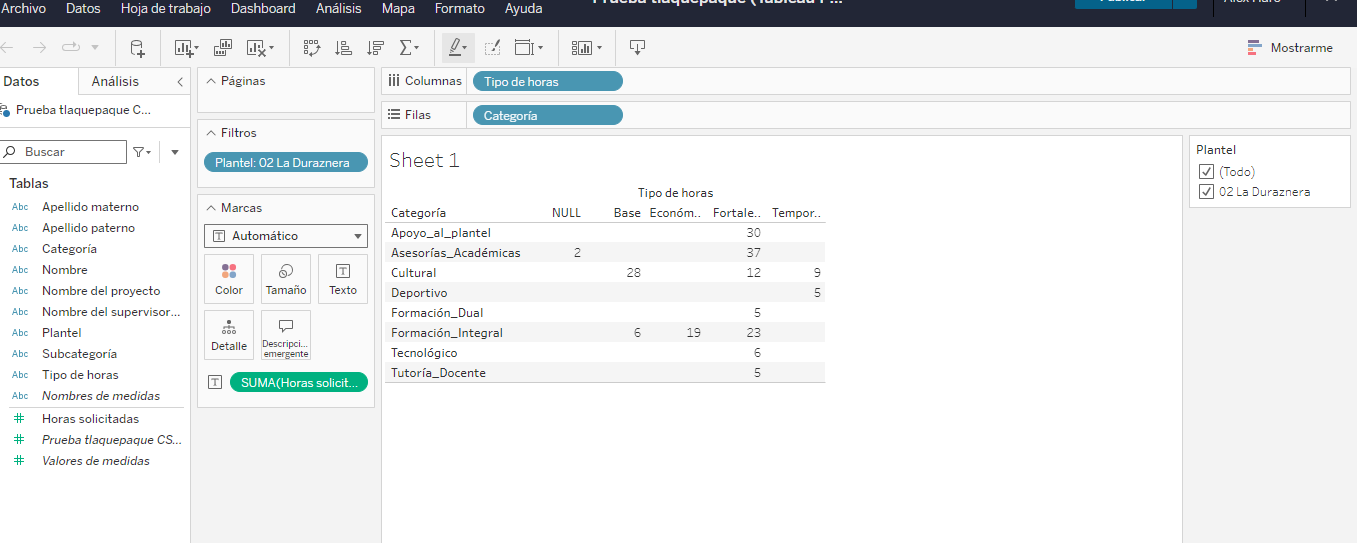
¿Quién es tu audiencia? ¿Cuál es la mejor manera de comunicarte con ellos?

Directores de área; con tablas y gráficas simples que reflejen la información recabada

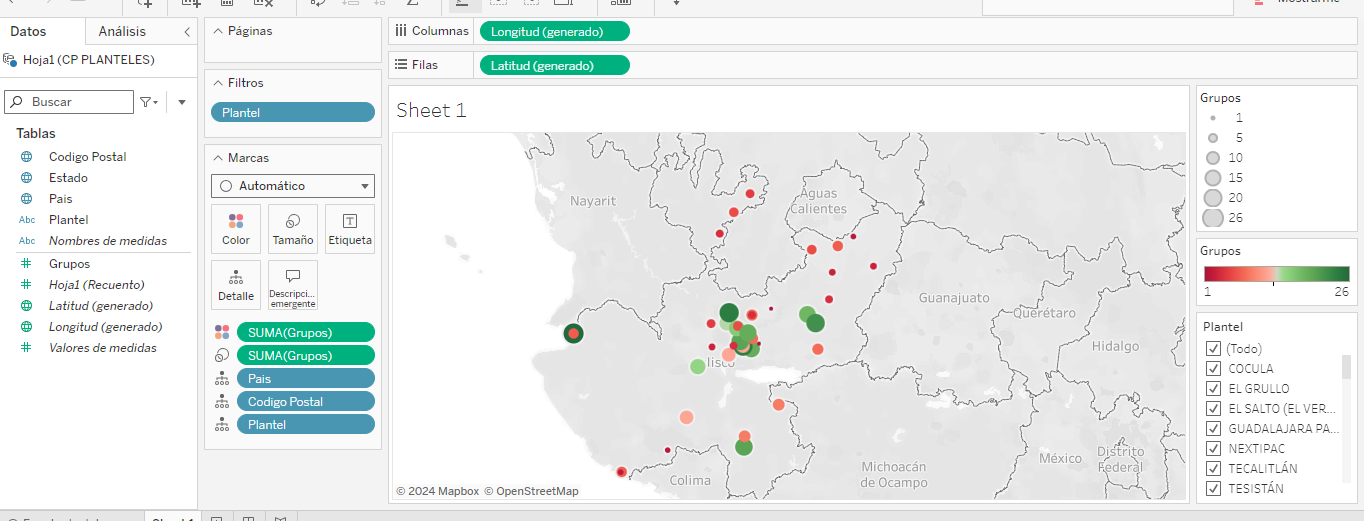
¿La visualización de datos puede ayudarte a compartir tus hallazgos?

Sí, de manera fundamental

**Resumen en tabla y geográfico en Tableu**

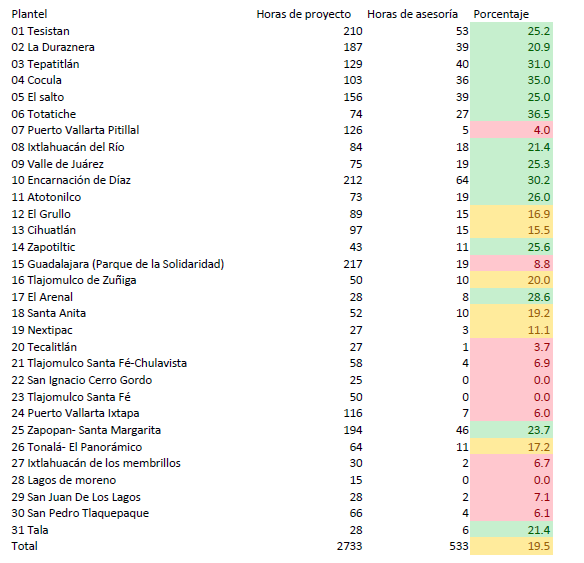


<https://public.tableau.com/app/profile/alex.haro5532/viz/ndicesdereprobacinFeb-Jul24/Sheet1>



<https://public.tableau.com/app/profile/alex.haro5532/viz/DistribucindehorasdeproyectoFeb-Jul24/Sheet1>

|  |  |  |  |
| --- | --- | --- | --- |
| **PLANTEL** | **Pais** | **Estado** | Codigo Postal |
| **TESISTÁN** | **Mexico** | **Jalisco** | 45200 |
| **LA DURAZNERA (TLAQUEPAQUE)** | **Mexico** | **Jalisco** | 45610 |
| **TEPATITLÁN** | **Mexico** | **Jalisco** | 47600 |
| **AULA EXTERNA SAN MIGUEL EL ALTO** | **Mexico** | **Jalisco** | 47140 |
| **COCULA** | **Mexico** | **Jalisco** | 48500 |
| **AULA EXTERNA AYOTITLAN** | **Mexico** | **Jalisco** | 48962 |
| **EL SALTO (EL VERDE)** | **Mexico** | **Jalisco** | 45685 |
| **AULA EXTERNA LA HIGUERA** | **Mexico** | **Jalisco** | 45696 |
| **TOTATICHE** | **Mexico** | **Jalisco** | 46170 |
| **AULA EXTERNA CHIMALTITAN** | **Mexico** | **Jalisco** | 46325 |
| **AULA EXTERNA COLOTLAN** | **Mexico** | **Jalisco** | 46200 |
| **PUERTO VALLARTA PITILLAL (LAS JUNTAS)** | **Mexico** | **Jalisco** | 48290 |
| **IXTLAHUACÁN DEL RÍO** | **Mexico** | **Jalisco** | 45260 |
| **AULA EXTERNA CUQUIO** | **Mexico** | **Jalisco** | 45480 |
| **VALLE DE JUÁREZ** | **Mexico** | **Jalisco** | 49540 |
| **ENCARNACIÓN DE DÍAZ** | **Mexico** | **Jalisco** | 47270 |
| **AULA EXTERNA BAJIO DE SAN JOSE** | **Mexico** | **Jalisco** | 47280 |
| **AULA EXTERNATEOCALTICHE** | **Mexico** | **Jalisco** | 47202 |
| **ATOTONILCO** | **Mexico** | **Jalisco** | 47750 |
| **EL GRULLO** | **Mexico** | **Jalisco** | 48760 |
| **AULA EXTERNA TONAYA** | **Mexico** | **Jalisco** | 48760 |
| **CIHUATLÁN** | **Mexico** | **Jalisco** | 48985 |
| **AULA EXTERNA LA HUERTA** | **Mexico** | **Jalisco** | 48980 |
| **ZAPOTILTIC** | **Mexico** | **Jalisco** | 49600 |
| **GUADALAJARA PARQUE SOLIDARIDAD** | **Mexico** | **Jalisco** | 44760 |
| **TLAJOMULCO DE ZÚÑIGA** | **Mexico** | **Jalisco** | 45640 |
| **EL ARENAL** | **Mexico** | **Jalisco** | 45350 |
| **SANTA ANITA** | **Mexico** | **Jalisco** | 45600 |
| **NEXTIPAC** | **Mexico** | **Jalisco** | 45220 |
| **TECALITLÁN** | **Mexico** | **Jalisco** | 49900 |
| **TLAJOMULCO SANTA FE -CHULAVISTA** | **Mexico** | **Jalisco** | 45655 |
| **SAN IGNACIO CERRO GORDO** | **Mexico** | **Jalisco** | 47190 |
| **TLAJOMULCO - SANTA FE** | **Mexico** | **Jalisco** | 45655 |
| **PUERTO VALLARTA - IXTAPA** | **Mexico** | **Jalisco** | 48290 |
| **ZAPOPAN - SANTA MARGARITA** | **Mexico** | **Jalisco** | 45140 |
| **AULA EXTERNA VALENTIN GOMEZ FARIAS** | **Mexico** | **Jalisco** | 45130 |
| **TONALÁ - EL PANORÁMICO** | **Mexico** | **Jalisco** | 45400 |
| **AULA EXTERNA TOLOLOTLAN** | **Mexico** | **Jalisco** | 45427 |
| **IXTLAHUACAN DE LOS MEMBRILLOS** | **Mexico** | **Jalisco** | 45260 |
| **LAGOS DE MORENO** | **Mexico** | **Jalisco** | 47443 |
| **SAN JUAN DE LOS LAGOS** | **Mexico** | **Jalisco** | 47030 |
| **SAN PEDRO TLAQUEPAQUE** | **Mexico** | **Jalisco** | 45610 |
| **TALA** | **Mexico** | **Jalisco** | 45303 |



6.- ACTUAR

¿Cuál es tu conclusión final en función de tu análisis?

¿Cómo puedes aplicar tus conclusiones?

¿Existen algunos pasos que tú o tus interesados puedan seguir según tus hallazgos?

¿Existen datos adicionales que podrías utilizar para ampliar tus hallazgos?

¿Cómo puedes presentar tu caso práctico en tu portfolio?