

Informe de Empleo Industrial y Desarrollo Sostenible en el mundo:

Análisis para la estrategia de ventas de una empresa

Tarea 1: Análisis de Datos Exploratorio (EDA):

Se seleccionaron dos bases de datos de gapminder.org para este ejercicio. Se realizó un análisis exhaustivo de estos conjuntos de datos, que consta de información sobre la tasa de empleo en el sector industrial l Índice de Desarrollo Sostenible (IDS) de diversos países en todo el mundo, segmentados anualmente.

Los datos incluyen información geográfica, es decir, el país y el continente, y temporal, es decir, el año. El rango de años cubiertos es desde 1991 hasta 2019, y la frecuencia de los datos es anual. Por lo tanto, la vigencia de los datos que tenemos no es actual ya los datos deben actualizarse anualmente y no tenemos datos actualizados desde 2019.

Se identificaron ciertas agrupaciones y relaciones entre los campos. Estas relaciones se visualizaron a través de una serie de gráficos, que ayudaron a destacar las tendencias y outliers.

De este análisis, se decidió incluir en la visualización final los campos de país, año, tasa de empleo y IDS.

```
datos <- read_excel("~/Modulo 7/industry_workers_percent_of_employment.xlsx")

#Preparación de datos a formato largo
datos_largos <- datos %>%
  pivot_longer(cols = -country, names_to = "year", values_to = "employment_rate")

#Análisis descriptivo
#Tasa de empleo global por año
global_employment_rate <- datos_largos %>%
  group_by(year) %>%
  summarise(mean_employment_rate = mean(employment_rate, na.rm = TRUE))

#Países que aumentaron su tasa de empleo:
employment_increase <- datos_largos %>%
  group_by(country) %>%
  arrange(year) %>%
  summarise(change_in_employment = employment_rate[n()] - employment_rate[1]) %>%
  filter(change_in_employment > 0)

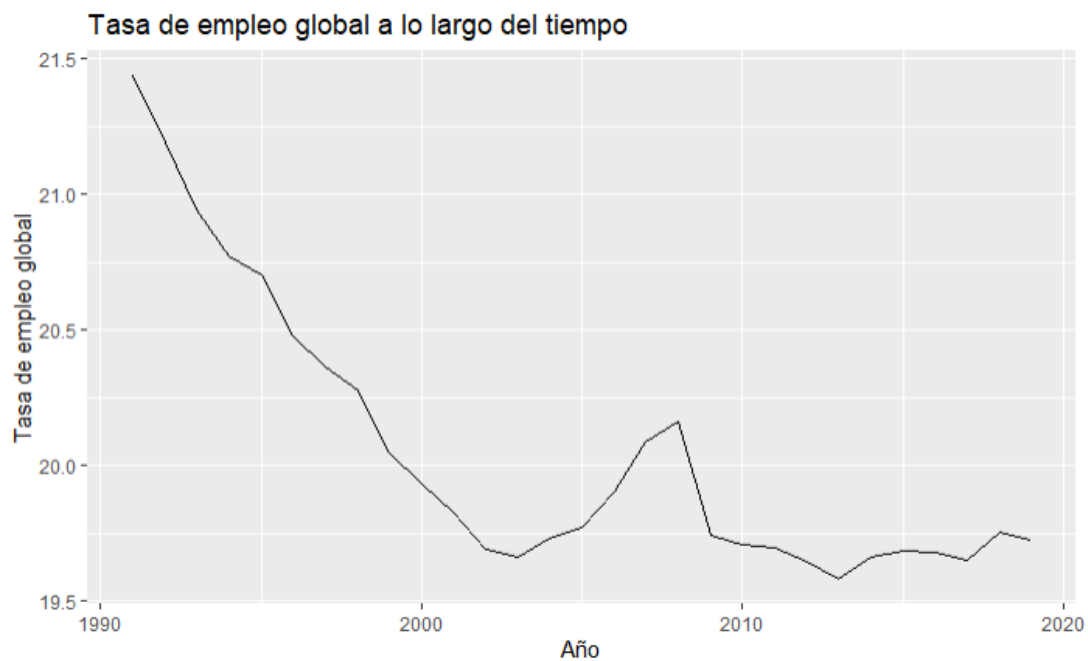
#Países que redujeron su tasa de empleo
employment_decrease <- datos_largos %>%
  group_by(country) %>%
  arrange(year) %>%
  summarise(change_in_employment = employment_rate[n()] - employment_rate[1]) %>%
  filter(change_in_employment < 0)
```

Calculamos la tasa de empleo global en el sector industrial a lo largo del tiempo:

```
# Visualización
# Convertir el año a numérico
global_employment_rate$year <- as.numeric(global_employment_rate$year)

# Crear la gráfica
ggplot(global_employment_rate, aes(x = year, y = mean_employment_rate, group=1)) +
  geom_line() +
  labs(title = "Tasa de empleo global a lo largo del tiempo",
       x = "Año", y = "Tasa de empleo global")
...

```



Definimos qué países pertenecen a cada continente y los asignamos y creamos la visualización:

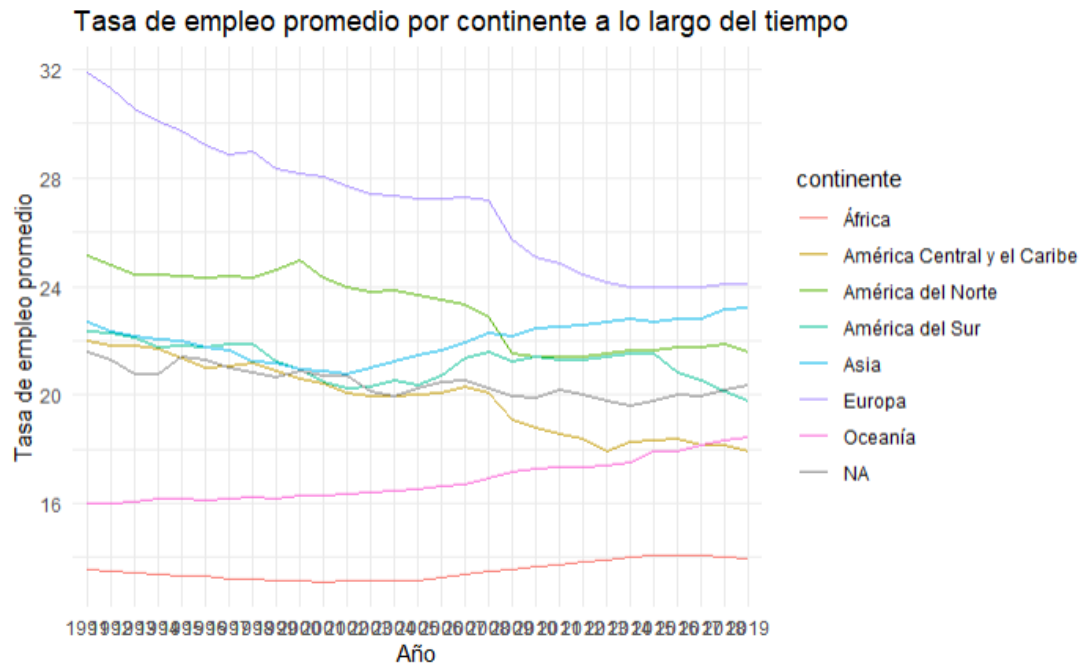
```
# Crear una nueva columna 'continente' con NA (sin asignación inicial)
datos$continente <- NA

# Luego asignar los continentes a cada país
datos_largos$continente[datos_largos$country %in% paises_africa] <- "África"
datos_largos$continente[datos_largos$country %in% paises_asia] <- "Asia"
datos_largos$continente[datos_largos$country %in% paises_europa] <- "Europa"
datos_largos$continente[datos_largos$country %in% paises_america_norte] <- "América del Norte"
datos_largos$continente[datos_largos$country %in% paises_america_sur] <- "América del Sur"
datos_largos$continente[datos_largos$country %in% paises_oceania] <- "Oceanía"
datos_largos$continente[datos_largos$country %in% paises_centro_caribe] <- "América Central y el Caribe"
...

```{r}
Calcular la media por continente y año
media_por_continente <- datos_largos %>%
 group_by(continente, year) %>%
 summarise(mean_employment_rate = mean(employment_rate, na.rm = TRUE), .groups = "drop")

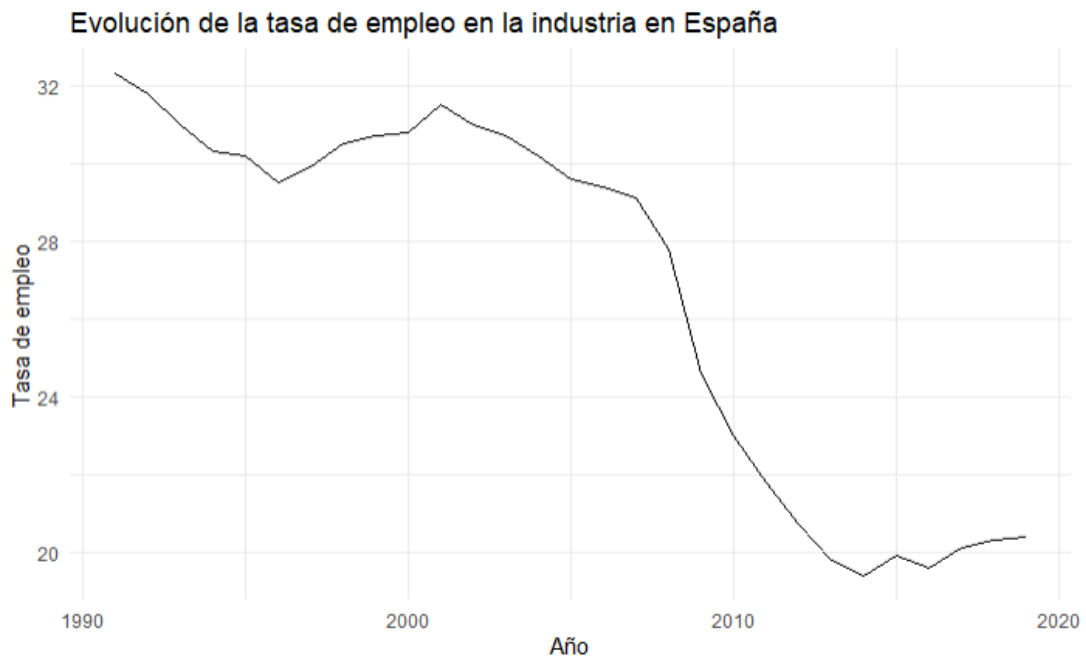
Crear el gráfico
ggplot(media_por_continente, aes(x = year, y = mean_employment_rate, color = continente)) +
 geom_line(aes(group = continente)) +
 labs(title = "Tasa de empleo promedio por continente a lo largo del tiempo",
 x = "Año", y = "Tasa de empleo promedio") +
 theme_minimal()

```



Calculamos la evolución de la tasa de empleo del sector industrial en España:

```
¿Cómo ha evolucionado la tasa de empleo en la industria en España a lo largo del tiempo?
ggplot(datos_spain, aes(x = year, y = employment_rate)) +
 geom_line() +
 labs(x = "Año", y = "Tasa de empleo") +
 theme_minimal() +
 ggtitle("Evolución de la tasa de empleo en la industria en España")
```



Calculamos la evolución de la tasa de empleo para España y Europa para ver una comparativa:

```

'''r
#Cálculos España
#¿Cómo se compara la tasa de empleo en la industria de España con la de otros países de la Unión Europea y del mundo?

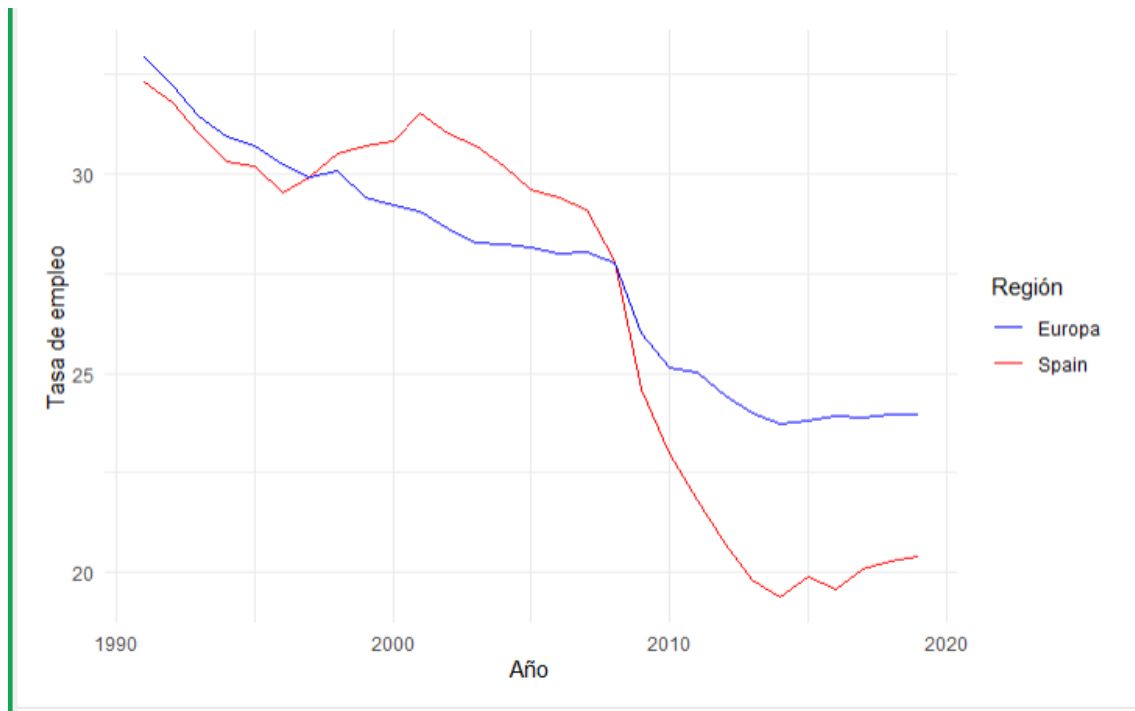
Datos de España
datos_spain <- datos_largos %>%
 filter(country == "Spain")

Datos de la Unión Europea
países_europa <- c("Austria", "Belgium", "Bulgaria", "Croatia", "Cyprus", "Czech Republic", "Denmark", "Estonia", "Finland", "France", "Germany", "Greece", "Hungary",
"Ireland", "Italy", "Latvia", "Lithuania", "Luxembourg", "Malta", "Netherlands", "Poland", "Portugal", "Romania", "Slovakia", "Slovenia", "Spain", "Sweden")
datos_europa <- datos_largos %>%
 filter(country %in% países_europa) %>%
 group_by(year) %>%
 summarise(mean_employment_rate = mean(employment_rate, na.rm = TRUE))

Convertir 'year' a numérico
datos_spain$year <- as.numeric(datos_spain$year)
datos_europa$year <- as.numeric(datos_europa$year)

Comparación
ggplot() +
 geom_line(data = datos_spain, aes(x = year, y = employment_rate, color = "Spain")) +
 geom_line(data = datos_europa, aes(x = year, y = mean_employment_rate, color = "Europa")) +
 labs(x = "Año", y = "Tasa de empleo", color = "Región") +
 theme_minimal() +
 scale_color_manual(values = c("Spain" = "red", "Europa" = "blue"))

```



Observamos qué países han experimentado un crecimiento significativo de su tasa de empleo en el sector industrial:

```

#¿Qué países han experimentado un crecimiento significativo en su tasa de empleo en el sector industrial?

Calcular la tasa de crecimiento
datos_largos$growth_rate <- datos_largos %>%
 group_by(country) %>%
 mutate(growth_rate = (employment_rate/lag(employment_rate) - 1) * 100)

Ordenar los países por la tasa de crecimiento promedio
países_crecimiento <- datos_largos %>%
 group_by(country) %>%
 summarise(mean_growth_rate = mean(employment_rate, na.rm = TRUE)) %>%
 arrange(desc(mean_growth_rate))

Mostrar los primeros 100 países
head(países_crecimiento, 100)

```

A tibble: 100 × 2

country <chr>	mean_growth_rate <dbl>
Qatar	44.04828
Czech Republic	40.02759
Taiwan	38.31724
Slovak Republic	38.19310
Slovenia	36.82414
Turkmenistan	36.27931
United Arab Emirates	34.32759
Bulgaria	33.88276
North Macedonia	32.87931
Mauritius	32.84828

1-10 of 100 rows

Añadimos a nuestro análisis los datos del dataset de Índice de Desarrollo Sostenible:

```
#Nuevo dataset Índice de desarrollo sostenible
Cargar el dataset
datos_sdi <- read_excel("sdi.xlsx")

Convertir el dataset a formato largo
datos_sdi_largos <- datos_sdi %>%
 gather(key = "year", value = "sdi", -country)
...

```{r}
# Unir los datos de los dos dataset
datos_unidos <- datos_largos %>%
  left_join(datos_sdi_largos, by = c("country", "year"))
...

```{r}
Calcular la correlación entre la tasa de empleo y el IDS
correlacion <- cor(datos_unidos$employment_rate, datos_unidos$sdi, use = "pairwise.complete.obs")

Imprimir el coeficiente de correlación
print(correlacion)
...

[1] 0.1890459
```

El Índice de Desarrollo Sostenible (IDS) puede proporcionar un marco integral para entender la salud económica, social y ambiental de un país, lo que es crucial para una empresa enfocada en apoyar los Objetivos de Desarrollo Sostenible (ODS). Un alto IDS puede indicar que un país ya está comprometido con la sostenibilidad y podría estar más abierto a productos y servicios que ayuden a avanzar en esa dirección.

Por otro lado, la tasa de empleo en el sector industria puede darte una idea de la escala y relevancia de tu mercado objetivo en un país en particular. Por ejemplo, si un país tiene una alta tasa de empleo en la industria, esto puede indicar una demanda potencialmente mayor de productos y servicios que apoyen la sostenibilidad en ese sector.

Finalmente, analizar la correlación entre el IDS y la tasa de empleo en la industria también puede revelar tendencias interesantes. Por ejemplo, si estos dos indicadores están fuertemente

correlacionados, esto podría sugerir que los esfuerzos de sostenibilidad pueden estar asociados con un sector industrial más saludable y, por lo tanto, un mayor mercado para tus productos.

En conclusión, estas variables definitivamente podrían proporcionar información valiosa para ayudarnos a tomar decisiones sobre la estrategia de ventas de la empresa.

En nuestro caso, un valor de 0.1890459 indica una correlación positiva débil entre las dos variables en cuestión. Esto probablemente se deba a otras variables y factores contextuales que no están reflejados en el gráfico. Por ejemplo, es posible que algunos países en este cuadrante ya estén tomando medidas para mejorar su sostenibilidad, o es posible que existan barreras culturales o regulatorias para adoptar nuevas tecnologías en ciertos países.

Para entender mejor la evolución y los datos, creamos un gráfico de dispersión con los datos de todos los países y años:

```
Unir los datos
datos_combinados <- merge(datos_largos, datos_sdi_largos, by = c("country", "year"), all = TRUE)

Gráfico de dispersión diferenciando los continentes
ggplot(datos_combinados, aes(x = employment_rate, y = sdi, color = continente)) +
 geom_point(na.rm = TRUE) +
 labs(x = "Tasa de empleo en la industria",
 y = "Índice de desarrollo sostenible",
 color = "Continente") +
 theme_minimal() +
 theme(legend.position = "bottom")
```



Observamos que en Asia, países como Japón, Singapur, Corea del Sur, y China han experimentado una industrialización significativa en las últimas décadas. Estos países tienen tasas de empleo en el sector industrial altas. En términos de IDS, Japón y Singapur generalmente tienen IDS altos debido a sus políticas ambientales, sociales y económicas. China e India han logrado avances significativos en ciertos aspectos del desarrollo sostenible, pero también enfrentan desafíos en áreas como la calidad del aire y la desigualdad de ingresos.

En África, el panorama es más variado. Países como Sudáfrica, Egipto y Nigeria tienen sectores industriales relativamente desarrollados en comparación con otras naciones africanas. En

términos de IDS, varios países africanos están realizando esfuerzos para lograr los Objetivos de Desarrollo Sostenible de las Naciones Unidas, aunque a menudo enfrentan desafíos debido a problemas como la pobreza, la inestabilidad política y los impactos del cambio climático.

En Europa, varios países tienen un sector industrial bien establecido y han hecho importantes progresos en términos de desarrollo sostenible.

Realizamos una regresión lineal:

```
#Rregresión lineal
modelo_lineal <- lm(sdi ~ employment_rate, data = datos_combinados)
summary(modelo_lineal)
```

```
Call:
lm(formula = sdi ~ employment_rate, data = datos_combinados)

Residuals:
 Min 1Q Median 3Q Max
-51.736 -9.262 2.807 13.017 29.536

Coefficients:
 Estimate Std. Error t value Pr(>|t|)
(Intercept) 49.16967 0.63750 77.13 <2e-16 ***
employment_rate 0.37101 0.02891 12.84 <2e-16 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 16.58 on 4445 degrees of freedom
(1256 observations deleted due to missingness)
Multiple R-squared: 0.03574, Adjusted R-squared: 0.03552
F-statistic: 164.7 on 1 and 4445 DF, p-value: < 2.2e-16
```

Coefficientes: Aquí vemos el resultado de la regresión. La estimación de la "intercepción" es el valor esperado de 'sdi' cuando 'employment\_rate' es 0. La estimación de 'employment\_rate' nos dice cuánto cambia el 'sdi' en promedio por cada aumento unitario en 'employment\_rate'.

Pr(>|t|): Esta es la probabilidad de que los coeficientes sean iguales a cero, dado los datos que tenemos. Si es muy pequeño (generalmente menor a 0.05), entonces podemos rechazar la hipótesis nula de que el coeficiente es igual a cero. En este caso, ambos coeficientes tienen un p-valor de menos de 2e-16, lo que sugiere que ambas relaciones son estadísticamente significativas.

R-cuadrado: El R-cuadrado es una medida de qué tan bien se ajusta el modelo a los datos. Un R-cuadrado de 1 indica que el modelo explica perfectamente toda la variación en 'sdi'. En este caso, el R-cuadrado es 0.03574, lo que indica que el modelo explica aproximadamente el 3.6% de la variabilidad en 'sdi'. Esto no es muy alto, lo que sugiere que hay muchos otros factores no incluidos en el modelo que afectan 'sdi'.

¿qué significa todo esto para tu estrategia de ventas? Aunque la tasa de empleo en la industria y el índice de desarrollo sostenible están significativamente relacionados, la tasa de empleo en la industria solo explica una pequeña parte de la variabilidad en el índice de desarrollo sostenible.

Esto sugiere que, aunque puede haber una relación entre estas dos variables, hay muchos otros factores en juego y tu empresa puede necesitar considerar otras variables al formular su estrategia de ventas. Por ejemplo, podrías considerar aspectos como el nivel de ingresos del país, la estabilidad política, la legislación medioambiental, etc.

Por tanto, deberemos escoger los países en los que estas otras variables sean favorables.

## **Tarea 2: Audiencia y Objetivo**

La audiencia objetivo de nuestro dashboard es el equipo directivo de ventas de una empresa cuyo negocio es la venta de productos que ayudan a otras empresas a conseguir los objetivos de desarrollo sostenible. Esta audiencia no tiene porqué tener conocimiento de este tipo de datos, ya que hasta ahora el crecimiento que hemos tenido ha sido nacional.

El informe busca ayudar a definir la estrategia de ventas de la empresa hacia empresas públicas y privadas de países que potencialmente podrían beneficiarse de nuestros productos y servicios para mejorar los Objetivos de Desarrollo Sostenible (ODS).

Las preguntas que este dashboard buscará responder incluyen:

- ¿Cómo ha evolucionado la tasa de empleo y el IDS en un país específico a lo largo de los años?
- ¿Cómo se comparan las tasas de empleo y los IDS entre diferentes países en un año específico?
- ¿Qué países tienen un bajo IDS y una alta tasa de empleo industrial?
- ¿Qué países son idóneos, por tanto, para enfocar nuestra estrategia de ventas en los próximos años?

Una vez decidida la estrategia, se podrían incorporar nuestros datos de ventas y analizar la correlación entre los países en los que hemos tenido más ventas y la evolución del índice de desarrollo sostenible.

## **Tarea 3: Selección de Gráficos y Codificación**

Para nuestra tarea de visualización, decidimos usar gráficos de líneas y gráficos de dispersión. La elección de estos tipos específicos de gráficos se basó en su funcionalidad, los objetivos que perseguíamos y los datos que íbamos a visualizar.

Los gráficos de líneas son particularmente eficaces para mostrar una serie de datos en un intervalo de tiempo, lo que nos permite visualizar tendencias y patrones a lo largo del tiempo. Esto se alinea con uno de nuestros objetivos principales, que es visualizar cómo han evolucionado la tasa de empleo y el Índice de Desarrollo Sostenible (IDS) en un país específico a lo largo de los años (WARE, 2013).

Los gráficos de dispersión son una excelente opción cuando queremos visualizar la relación entre dos variables cuantitativas, como es nuestro caso con la tasa de empleo y el IDS. Este tipo de gráfico nos permite identificar correlaciones, agrupaciones y outliers entre los países.



Además, al agregar una dimensión de color, podemos distinguir entre diferentes categorías (continentes, en nuestro caso), que enriquecen la información disponible en el gráfico (YAU, 2011).

La codificación de las variables fue elegida con un énfasis en la simplicidad y la claridad. Se utilizó una escala lineal para la tasa de empleo y el IDS, ya que estos datos son cuantitativos y continuos. Estas decisiones están en línea con los principios de la teoría de la percepción y las mejores prácticas de diseño de visualización de datos (FEW, 2013).

La codificación de color se utilizó para representar las diferentes categorías (países), lo que permite una identificación visual rápida y fácil. La elección del color se realizó teniendo en cuenta los principios de la percepción y las recomendaciones de las referencias WARE, 2013 y YAU, 2011.

A continuación, detallo cómo las decisiones de análisis y diseño se alinean con las recomendaciones y mejores prácticas:

- **Eficacia de los gráficos de líneas:** Ware (2013) menciona que los gráficos de líneas son una de las formas más efectivas de representar cambios continuos en el tiempo debido a la naturaleza de nuestra percepción visual. Nuestro sistema visual es muy bueno para seguir líneas y percibir tendencias y patrones en los datos de tiempo. Esto justifica nuestra elección de los gráficos de líneas para representar la evolución del IDS y la tasa de empleo a lo largo del tiempo.
- **Representación precisa de la magnitud:** Los gráficos que elegimos, tanto de líneas como de dispersión, nos permiten representar la magnitud de las variables (IDS y tasa de empleo) de una manera precisa. Ware (2013) indica que la longitud y la posición a lo largo de un eje común son las formas más precisas de representar cantidades. En nuestros gráficos, la tasa de empleo y el IDS se representan a lo largo de ejes comunes, proporcionando una visualización precisa.
- **Distinguibilidad de colores:** Para diferenciar los continentes en nuestro gráfico de dispersión, utilizamos la codificación de colores. Ware (2013) y Yau (2011) mencionan que nuestro sistema visual puede distinguir fácilmente entre diferentes colores, lo que nos permite utilizar el color para representar categorías diferentes. En nuestro caso, usamos colores para representar diferentes continentes, lo que facilita la comparación entre países de diferentes continentes.
- **Uso de herramientas interactivas:** Nuestra elección de utilizar una aplicación Shiny también se basa en los principios de percepción y en las mejores prácticas de diseño de visualización de datos. La interactividad permite a los usuarios explorar los datos a su propio ritmo y según sus propios intereses, lo que mejora la comprensión y la participación (Yau, 2011).
- **Facilitar la comparación:** Al proporcionar dos gráficos de líneas en la misma pantalla, permitimos una comparación directa entre dos países. Esta disposición se alinea con el principio de proximidad de la teoría de la Gestalt, donde los elementos que están cerca unos de otros se perciben como un grupo. Esto facilita la comparación directa y la identificación de diferencias y similitudes (Ware, 2013).

#### **Tarea 4: Disposición e Interacción**

Tenemos dos aspectos clave del diseño de un dashboard: la disposición de los gráficos y la interacción. Ambos aspectos están profundamente arraigados en los principios de la psicología de la percepción, la interacción del usuario y la experiencia del usuario.

Justificación de la disposición de los gráficos en el Dashboard y la interacción:

**Disposición de los gráficos:** En nuestro dashboard, optamos por una disposición vertical de los gráficos de línea y un diseño de tabulación para alternar entre los diferentes tipos de gráficos (línea, dispersión y tabla). La disposición vertical de los gráficos de líneas facilita la comparación visual directa de las tendencias entre dos países diferentes (principio de proximidad de la teoría de la Gestalt mencionada por Ware (2013)). Al mismo tiempo, el diseño de tabulación nos permite mantener el dashboard limpio y sin saturación visual, que es una práctica recomendada en el diseño de dashboards (Few, 2013).

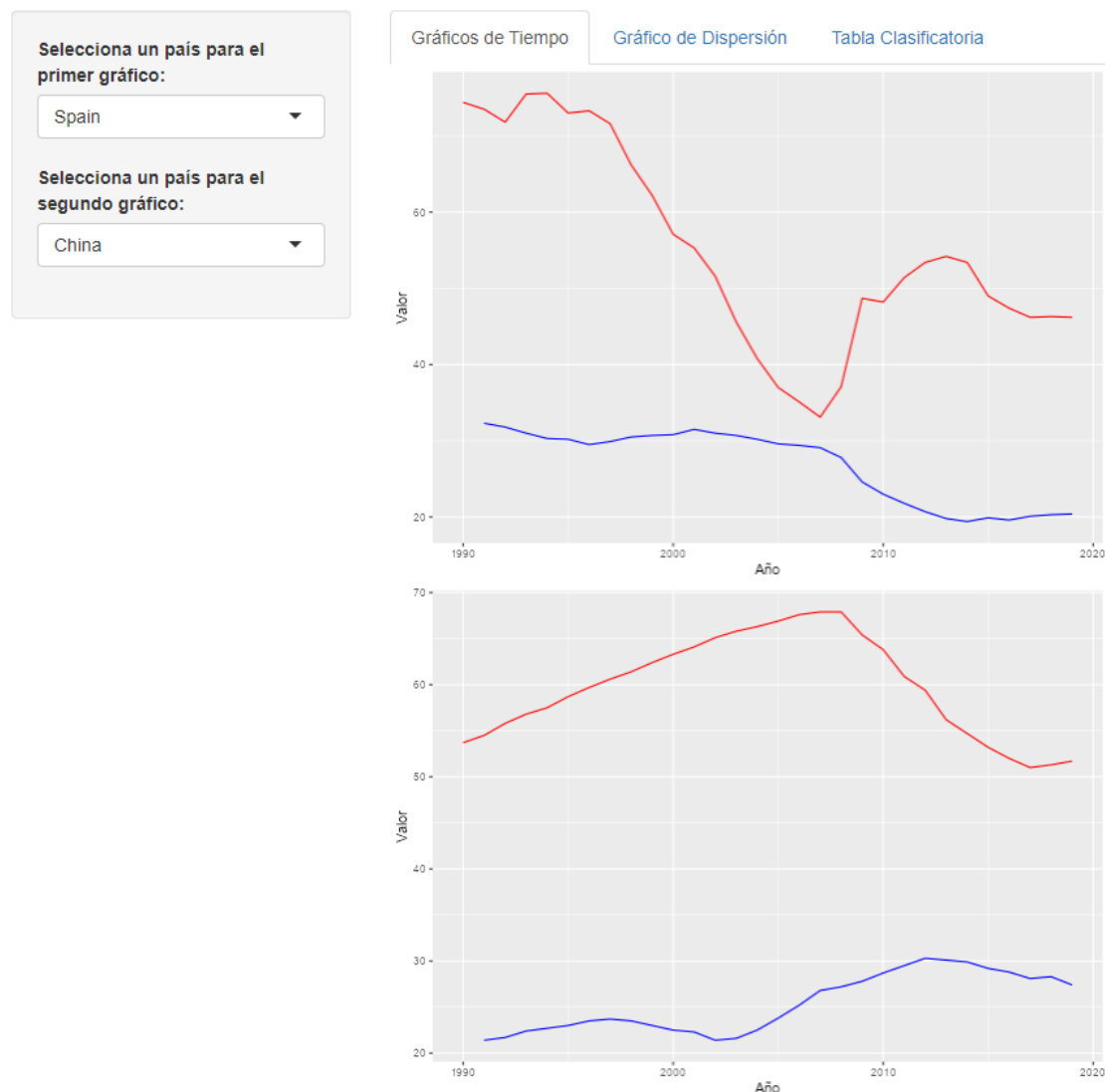
- La "*disposición vertical de los gráficos de línea*" se refiere a cómo colocamos los dos gráficos de línea en nuestra interfaz de usuario. En lugar de colocarlos uno al lado del otro (horizontalmente), los colocamos uno encima del otro (verticalmente). Esta elección permite a los usuarios comparar más fácilmente las tendencias a lo largo del tiempo entre los dos países seleccionados, ya que sus ojos pueden moverse hacia arriba y hacia abajo más fácilmente que de izquierda a derecha cuando se trata de gráficos de tiempo.
- El "*diseño de tabulación para alternar entre los diferentes tipos de gráficos*" se refiere a la forma en que organizamos los diferentes gráficos en nuestro dashboard. En lugar de mostrar todos los gráficos (líneas, dispersión, tabla) a la vez, lo que podría saturar visualmente a los usuarios y hacer que sea difícil centrarse en una pieza de información a la vez, los colocamos en pestañas separadas. Los usuarios pueden entonces hacer clic en la pestaña correspondiente ("Gráficos de Tiempo", "Gráfico de Dispersión", "Tabla Clasificatoria") para ver el tipo de gráfico que les interesa en ese momento. Esta elección ayuda a mantener nuestro dashboard limpio y fácil de usar, al tiempo que permite a los usuarios explorar los datos en profundidad a su propio ritmo.

**Interacción:** Nuestro dashboard es interactivo, permitiendo a los usuarios seleccionar los países que desean comparar, así como el rango de años que desean visualizar. Este nivel de interactividad mejora la utilidad y usabilidad del dashboard, ya que permite a los usuarios explorar los datos en profundidad y de acuerdo a sus necesidades particulares (Ware, 2013; Few, 2013). Además, la implementación de tooltips en el gráfico de dispersión permite a los usuarios obtener más detalles sobre cada punto, lo cual es particularmente útil dada la cantidad de datos que se representan en este gráfico.

Ambos aspectos están diseñados para trabajar juntos para mejorar la utilidad, la facilidad de uso y la eficiencia del dashboard. Las decisiones específicas que he tomado están en línea con las mejores prácticas y principios destacados por Ware (2013) y Few (2013), y se basan en la idea de hacer que los datos sean accesibles, comprensibles y útiles para nuestra audiencia objetivo.

## Tarea 5: Bocetos iniciales

# Dashboard de Empleo Industrial y Desarrollo Sostenible



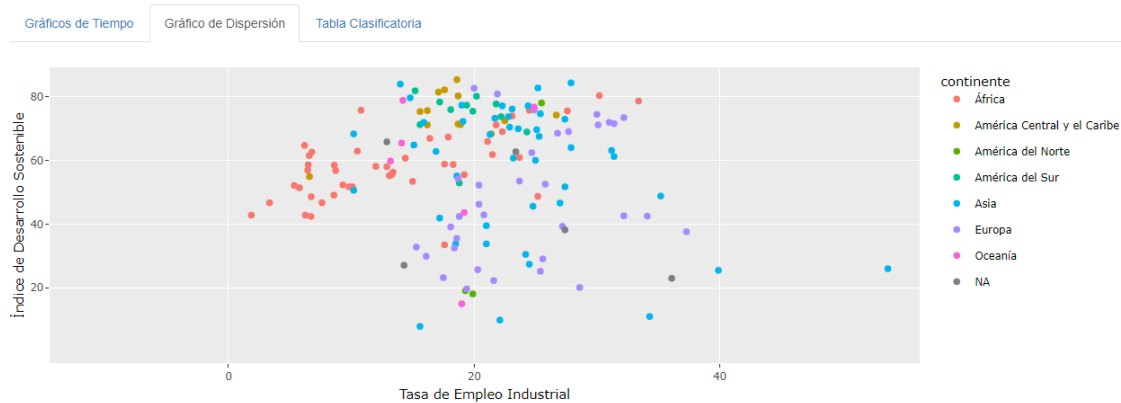
Los gráficos de tiempo en nuestro dashboard muestran la evolución del Índice de Desarrollo Sostenible (SDI) y la tasa de empleo en el sector industrial en un país seleccionado durante un periodo de tiempo especificado por el usuario.

Cada gráfico de tiempo consiste en dos líneas, cada una de las cuales representa una de las dos variables mencionadas (SDI y tasa de empleo industrial). El eje horizontal (eje x) representa el tiempo, específicamente los años, y el eje vertical (eje y) representa los valores de SDI y tasa de empleo.

La línea azul en cada gráfico representa la tasa de empleo en el sector industrial. Esto proporciona una visión clara de cómo ha cambiado la tasa de empleo en el sector industrial a lo largo del tiempo. Un aumento en esta línea indica un aumento en la tasa de empleo, mientras que una disminución en la línea indica una disminución en la tasa de empleo.

Por otro lado, la línea roja en cada gráfico representa el SDI. Esto permite al usuario ver cómo ha evolucionado el SDI del país seleccionado a lo largo del tiempo. Al igual que con la línea azul, un aumento en la línea roja indica un aumento en el SDI, mientras que una disminución indica una disminución en el SDI.

Además, dado que hay dos gráficos de tiempo en el dashboard, los usuarios pueden seleccionar diferentes países para cada gráfico. Esto permite comparaciones directas entre los dos países en términos de cómo han cambiado tanto el SDI como la tasa de empleo en el sector industrial a lo largo del tiempo.



El gráfico de dispersión en nuestro dashboard proporciona una representación visual de la relación entre el Índice de Desarrollo Sostenible (SDI) y la tasa de empleo en el sector industrial en varios países en un año determinado, que en este caso hemos fijado en 2019.

En este gráfico, el eje horizontal (eje x) representa la tasa de empleo en el sector industrial, y el eje vertical (eje y) representa el SDI. Cada punto en el gráfico representa un país, y la posición de cada punto está determinada por los valores de estas dos variables para ese país.

Además, cada punto está codificado por color para representar el continente al que pertenece el país. Esto permite al usuario identificar rápidamente agrupaciones geográficas y hacer comparaciones entre países de diferentes continentes.

Cuando el usuario pasa el ratón sobre un punto, se muestra un tooltip con el nombre del país y su continente, lo que proporciona un contexto adicional.

El gráfico de dispersión permite a los usuarios identificar tendencias y patrones en los datos, así como posibles outliers. Por ejemplo, los países con un SDI alto y una tasa de empleo en el sector industrial baja se agruparán en la parte superior izquierda del gráfico, mientras que los países con un SDI bajo y una alta tasa de empleo en el sector industrial se agruparán en la parte inferior derecha.

```

{r}

Calcular los percentiles para sdi y employment_rate
sdi_threshold <- quantile(datos_combinados$sdi, 0.25, na.rm = TRUE)
employment_rate_threshold <- quantile(datos_combinados$employment_rate, 0.75, na.rm = TRUE)

Filtrar los datos para los países con bajo sdi y alta tasa de empleo industrial
datos_filtrados <- datos_combinados %>%
 filter(sdi <= sdi_threshold & employment_rate >= employment_rate_threshold)

Obtener la lista de países
países_potenciales <- unique(datos_filtrados$country)

Imprimir la lista de países
print(países_potenciales)

```

Finalmente, con este código realizamos una operación de selección de datos para identificar países que podrían ser de interés para ciertas estrategias de negocio o desarrollo. Más específicamente, está buscando países que tengan un Índice de Desarrollo Sostenible relativamente bajo y una alta tasa de empleo en el sector industrial. La lógica detrás de esto sería que estos países estarán más dispuestos a invertir en soluciones para mejorar su SDI, mientras que al mismo tiempo tienen una industria fuerte que podría beneficiarse de tales soluciones.

En primer lugar, calculamos los umbrales para lo que considera "bajo" en términos de SDI y "alto" en términos de tasa de empleo en el sector industrial. Esto se hace a través del cálculo de los percentiles. El umbral de SDI se establece en el percentil 25, lo que significa que aquellos países cuyo SDI esté en el cuarto inferior se considerarán como de "bajo" SDI. De manera similar, el umbral de la tasa de empleo en el sector industrial se establece en el percentil 75, es decir, los países en el cuarto superior se considerarán como de "alta" tasa de empleo en el sector industrial.

A continuación, filtra el conjunto de datos para seleccionar solo aquellos países que cumplen con ambos criterios: SDI bajo y alta tasa de empleo en el sector industrial.

Finalmente, genera una lista de estos países, removiendo cualquier duplicado, y los imprime para su visualización. Esta lista de países podría considerarse como potencialmente beneficiosa para enfocar las estrategias de negocio o desarrollo relacionadas con la mejora del SDI.

```

Imprimir la lista de países
print(países_potenciales)

```

[1] "Austria"	"Bahrain"	"Burkina Faso"	"Czech Republic"
[5] "Estonia"	"Finland"	"Germany"	"Iceland"
[9] "Ireland"	"Italy"	"Japan"	"Lithuania"
[13] "Malta"	"Poland"	"Qatar"	"Serbia"
[17] "Singapore"	"Slovak Republic"	"Slovenia"	"South Korea"
[21] "Spain"	"Turkmenistan"	"United Arab Emirates"	"United States"

Por último, deberíamos realizar un análisis de los factores que pueden favorecer que nuestro producto sea comprado en esos países y ya tendríamos toda la información necesaria para definir nuestra estrategia de ventas para la expansión de nuestra empresa.

A continuación se encuentra el código para la visualización del dashboard:

```
datos_combinados$year <- as.numeric(datos_combinados$year)

ui <- fluidPage(
 titlePanel("Dashboard de Empleo Industrial y Desarrollo Sostenible"),
 sidebarLayout(
 sidebarPanel(
 selectInput("countries1", "Selecciona un país para el primer gráfico:",
 choices = unique(datos_combinados$country)),
 selectInput("countries2", "Selecciona un país para el segundo gráfico:",
 choices = unique(datos_combinados$country))
),
 mainPanel(
 tabsetPanel(
 tabPanel("Gráficos de Tiempo",
 plotOutput("linePlot1"),
 plotOutput("linePlot2")),
 tabPanel("Gráfico de Dispersión", plotlyOutput("scatterPlot")),
 tabPanel("Tabla Clasificatoria", tableOutput("rankTable"))
)
)
)
)

server <- function(input, output) {
 output$linePlot1 <- renderPlot({
 req(input$countries1)
 datos_seleccionados <- subset(datos_combinados, country == input$countries1)
 ggplot(datos_seleccionados, aes(x = year)) +
 geom_line(aes(y = employment_rate), color = "blue") +
 geom_line(aes(y = sdi), color = "red") +
 labs(x = "Año", y = "Valor", color = "Indicador")
 })

 output$linePlot2 <- renderPlot({
 req(input$countries2)
 datos_seleccionados <- subset(datos_combinados, country == input$countries2)
 ggplot(datos_seleccionados, aes(x = year)) +
 geom_line(aes(y = employment_rate), color = "blue") +
 geom_line(aes(y = sdi), color = "red") +
 labs(x = "Año", y = "Valor", color = "Indicador")
 })

 output$scatterPlot <- renderPlotly({
 datos_seleccionados <- subset(datos_combinados, year == 2019)
 plot <- ggplot(datos_seleccionados, aes(x = employment_rate, y = sdi, color = continente, text = paste("País: ", country, "
 Continente: ", continente))) +
 geom_point() +
 labs(x = "Tasa de Empleo Industrial", y = "Índice de Desarrollo Sostenible")
 ggplotly(plot, tooltip = "text")
 })

 output$rankTable <- renderTable({
 datos_seleccionados <- subset(datos_combinados, year == 2019)
 tabla_rank <- datos_seleccionados %>%
 arrange(desc(sdi), desc(employment_rate)) %>%
 select(country, year, sdi, employment_rate)
 })
}

shinyApp(ui = ui, server = server)
```

Aspectos importantes a tener en cuenta:

**Economía:** Los países con economías fuertes y desarrolladas, como Alemania y Estados Unidos, podrían tener más empresas que estén dispuestas y sean capaces de invertir en productos que ayuden a lograr los ODS.

**Políticas de sostenibilidad:** Algunos países tienen políticas gubernamentales fuertes de apoyo a los ODS. Por ejemplo, Finlandia es reconocido por su compromiso con la sostenibilidad y tiene una serie de políticas en vigor para fomentar la sostenibilidad en las empresas.

**Industrias existentes:** Algunos países tienen industrias fuertes que son relevantes para los ODS. Por ejemplo, Austria tiene una industria de energías renovables en crecimiento y puede ser un objetivo potencial para productos relacionados con la energía sostenible.

**Interés en la sostenibilidad:** Algunos países, como Islandia, tienen un alto interés en la sostenibilidad, lo que podría significar que hay una mayor demanda de productos que ayuden a alcanzar los ODS.

Crecimiento económico y potencial de desarrollo: Los países en desarrollo o con economías emergentes, como Qatar y Emiratos Árabes Unidos, pueden estar buscando tecnologías y soluciones sostenibles para apoyar su crecimiento económico.

Teniendo en cuenta estos factores, recomendaría centrar nuestra estrategia de ventas en estos cinco países:

- Alemania
- Estados Unidos
- Finlandia
- Austria
- Islandia