

# **Universidad Torcuato di Tella**

## **Master in Management + Analytics**

**Ventas y Marketing: Trabajo Final**  
Richard Rivas

**Análisis de Conjuntos para Paquetes Turísticos**

## Diseño del Estudio y Recolección de Datos

Para este estudio se seleccionó el producto de paquetes turísticos para conocer las diferentes preferencias de los consumidores en cuanto a destinos, tiempo de estadía y comodidades por las que estarían dispuestos a pagar.

Se consideraron 4 atributos clave en la oferta de los paquetes turísticos, a continuación se presentan sus categorías y niveles:

**1) Destino turístico (Lugar):** "Bariloche", "Salta", "Mendoza", "El Chaltén" .

**2) Tipo de Hotel (Hotel):**

- **Básico:** Internet y TV con cable, **sin** pileta o gimnasio, **sin** desayuno incluido.
- **Standard:** Internet y TV con cable, **sin** pileta o gimnasio, **con** desayuno incluido.
- **De Lujo:** Internet, y TV con cable, **con** pileta y gimnasio, **con** desayuno incluido.

**3) Precio (en ARS):** \$60.000, \$90.000, \$120.000

**4) Número de días (Días):** 3, 4 o 5

A partir de estas definiciones se creó la encuesta para la evaluación de cada perfil utilizando el diseño experimental ortogonal provisto por el código en R de la consigna de este trabajo. Posteriormente se envió la encuesta a amigos y familiares para realizar la recolección de datos.

Puede revisar la encuesta en [este enlace](#).

## Análisis de Datos Obtenidos

### 1. Realice una regresión lineal utilizando R o Excel para estimar los valores parciales de cada uno de sus encuestados.

Se realizó una regresión lineal para cada uno de los encuestados. En la imagen puede observarse la salida en R de uno de ellos tomado al azar con sus valores parciales asociados.

Para el intercepto, se tomaron las siguientes atributos:

- **Hotel:** De Lujo
- **Precio:** \$120.000
- **Lugar:** El Chaltén
- **Días:** 5

El resto de las regresiones pueden observarse en el archivo de R Markdown presente en el archivo comprimido de esta entrega.

A continuación puede observarse la fórmula utilizada para la regresión (realizada en un ciclo for, siendo i el ID del encuestado):

```
[1] "_____"
```

```
[1] "Regresion encuestado: 1"
```

```
[1] "_____"
```

```
Call:
lm(formula = Preference ~ Lugar_Bariloche + Lugar_Salta + Lugar_Mendoza +
    Hotel_Basico + Hotel_Standard + Precio_60k + Precio_90k +
    Dias_3 + Dias_4, data = filter(df, ID == i))
```

```
Residuals:
```

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
	0.1081	-0.8423	0.9144	0.1081	-0.2883	0.7117	0.4279	0.2973	-0.6486	-0.7883	-0.4054	-0.1396	-0.4234	0.5405	0.4279

```
Coefficients:
```

	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	7.4054	0.7734	9.576	0.00021 ***
Lugar_Bariloche	0.1667	0.6566	0.254	0.80972
Lugar_Salta	-2.0000	0.7581	-2.638	0.04608 *
Lugar_Mendoza	0.6667	0.7581	0.879	0.41946
Hotel_Basico	-2.7297	0.6106	-4.471	0.00658 **
Hotel_Standard	-1.9459	0.6106	-3.187	0.02435 *
Precio_60k	3.0000	0.6106	4.913	0.00442 **
Precio_90k	1.2973	0.6106	2.125	0.08699 .
Dias_3	-1.7838	0.6106	-2.921	0.03296 *
Dias_4	-1.0541	0.6106	-1.726	0.14489

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

Residual standard error: 0.9285 on 5 degrees of freedom
Multiple R-squared:  0.9278,    Adjusted R-squared:  0.7979 
F-statistic: 7.143 on 9 and 5 DF,  p-value: 0.02168
```

```
#Creamos las regresiones para cada encuestado
reg = lm(Preference ~ Lugar_Bariloche + Lugar_Salta + Lugar_Mendoza + Hotel_Basico + Hotel_Standard + Precio_60k + Precio_90k + Dias_3 + Dias_4, filter(df, ID == i))
```

## Análisis de Datos Obtenidos

### 1. Realice una regresión lineal utilizando R o Excel para estimar los valores parciales de cada uno de sus encuestados.

Tomando aleatoriamente un participante, puede observarse una preferencia marcada por el paquete turístico con más comodidades y días con destino a El Chaltén.

Sin embargo, al observar el valor parcial del precio en \$60.000 con 3 puntos, puede concluirse en el usuario valoraría una disminución en el precio, esta es una tendencia que puede notarse en la mayoría de los encuestados.

También vale la pena resaltar que Salta no sería el destino preferido de esta persona, representando un valor parcial de -2.

```
[1] "_____"
```

```
[1] "Regresion encuestado: 1"
```

```
[1] "_____"
```

```
Call:
lm(formula = Preference ~ Lugar_Bariloche + Lugar_Salta + Lugar_Mendoza +
    Hotel_Basico + Hotel_Standard + Precio_60k + Precio_90k +
    Dias_3 + Dias_4, data = filter(df, ID == i))
```

```
Residuals:
```

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
0.1081	-0.8423	0.9144	0.1081	-0.2883	0.7117	0.4279	0.2973	-0.6486	-0.7883	-0.4054	-0.1396	-0.4234	0.5405	0.4279

```
Coefficients:
```

	Estimate	Std. Error	t value	Pr(> t )	
(Intercept)	7.4054	0.7734	9.576	0.00021	***
Lugar_Bariloche	0.1667	0.6566	0.254	0.80972	
Lugar_Salta	-2.0000	0.7581	-2.638	0.04608	*
Lugar_Mendoza	0.6667	0.7581	0.879	0.41946	
Hotel_Basico	-2.7297	0.6106	-4.471	0.00658	**
Hotel_Standard	-1.9459	0.6106	-3.187	0.02435	*
Precio_60k	3.0000	0.6106	4.913	0.00442	**
Precio_90k	1.2973	0.6106	2.125	0.08699	.
Dias_3	-1.7838	0.6106	-2.921	0.03296	*
Dias_4	-1.0541	0.6106	-1.726	0.14489	

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

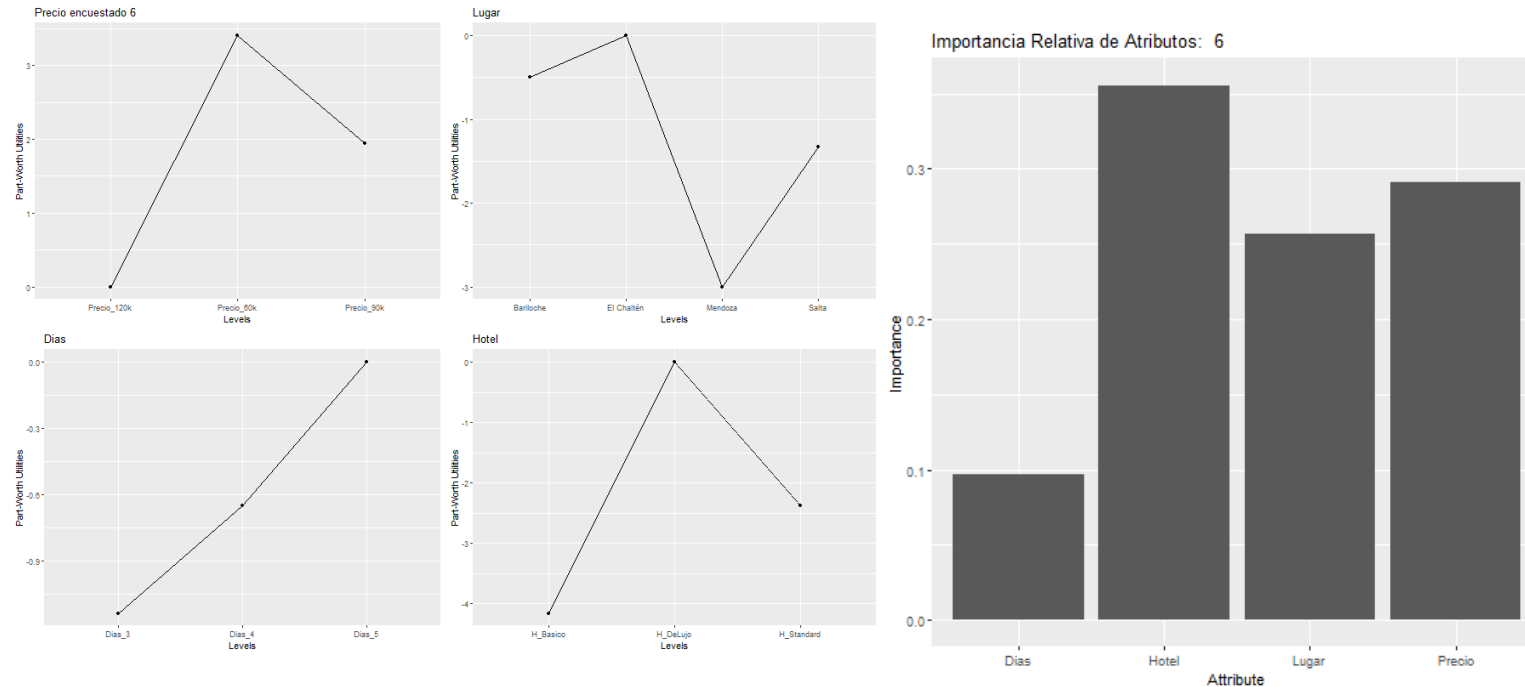
Residual standard error: 0.9285 on 5 degrees of freedom
Multiple R-squared:  0.9278,    Adjusted R-squared:  0.7979 
F-statistic: 7.143 on 9 and 5 DF,  p-value: 0.02168
```

```
#Creamos las regresiones para cada encuestado
reg = lm(Preference ~ Lugar_Bariloche + Lugar_Salta + Lugar_Mendoza + Hotel_Basico + Hotel_Standard + Precio_60k + Precio_90k + Dias_3 + Dias_4, filter(df, ID == i))
```

## Análisis de Datos Obtenidos

### 2. Para cada encuestado, determine la importancia relativa de cada atributo.

Se realizaron gráficas con los valores parciales de cada encuestado, así como de sus valores relativos, a continuación se observan los correspondientes al encuestado número 6, tomado al azar. Pueden observarse todos los gráficos en la sección de Anexos.



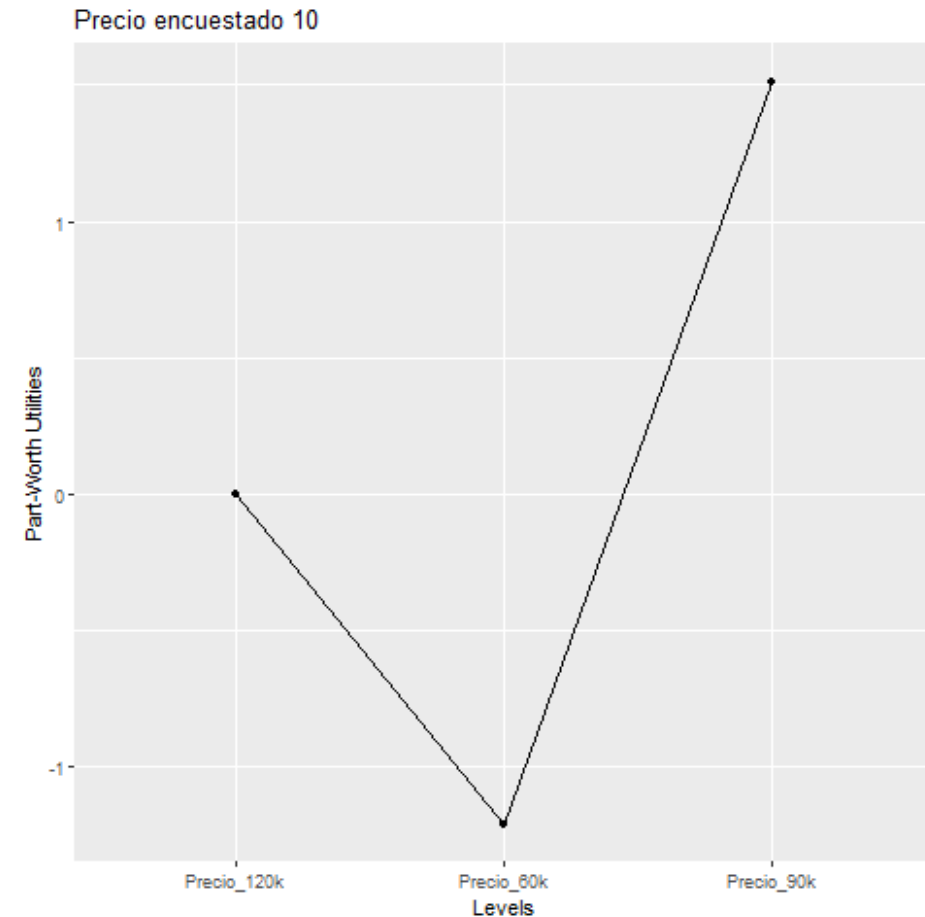
Puede observarse que el Hotel es el atributo más importante para este encuestado, seguido por el precio, lugar, y días. Siendo un hotel básico el atributo que le reportaría la menor utilidad, y prefiriendo sobre todos los demás lugares El Chaltén, por una estadía de 5 días.

## Análisis de Datos Obtenidos

### 3. Grafique los valores parciales asociados al precio de cada encuestado. ¿Cuáles son las implicancias de estos gráficos en las decisiones de gestión?

Estos gráficos permiten conocer la valoración de los clientes del producto en relación a su precio, lo cual, junto con los cálculos relacionados al monto que los clientes están dispuestos a pagar con respecto a un producto o servicio, puede realizarse un análisis que habilite a los tomadores de decisiones a optar por una venta masiva con bajo precio o reducida con un precio mayor.

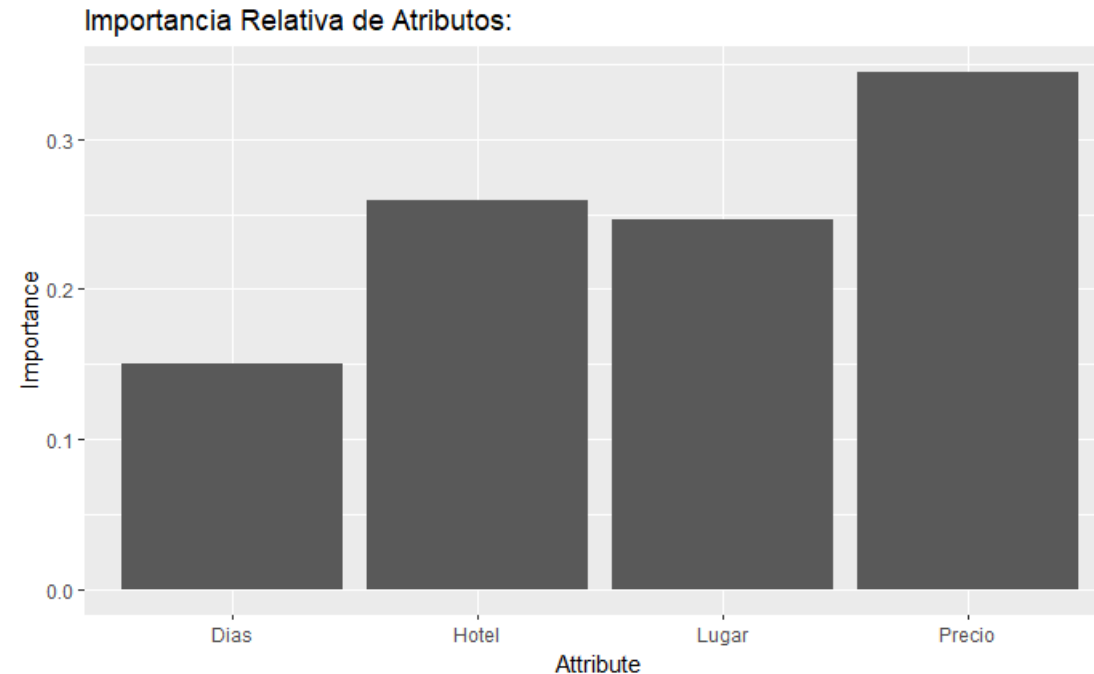
En este caso, para el encuestado número 10, se observa que de entrada podría pagar \$120.000 por un paquete turístico bastante completo, pero si el precio del mismo paquete fuera de \$60.000, probablemente podría generarle desconfianza, ya que le estaría restando utilidad. Sin embargo, el precio que más se ajusta a las preferencias de este cliente sería el de \$90.000. Este tipo de análisis, al complementarse con una segmentación adecuada de clientes, puede favorecer el diseño de diferentes paquetes que se ajusten correctamente a sus necesidades.



## Análisis de Datos Obtenidos

### 4. ¿Cuánto están dispuestos a pagar los consumidores por cambios en el atributo más importante?

Se realizó un análisis general del dataset para determinar el atributo más importante para los clientes y se determinó que este sería el precio, por lo que se procedió a realizar el análisis de disposición a pagar con el atributo de Hotel. Posteriormente se realizó el análisis para todos los encuestados.



## Análisis de Datos Obtenidos

### 4. ¿Cuánto están dispuestos a pagar los consumidores por cambios en el atributo más importante?

Para calcular el *willingness to pay*, primero se restó la utilidad del atributo del intercepto (Hotel de Lujo) con la utilidad del Hotel Básico:

$$\text{Rango\_uHotel} = u(\text{Hotel de Lujo}) - u(\text{Hotel Básico})$$

Posteriormente se calculó el rango de precio y su utilidad:

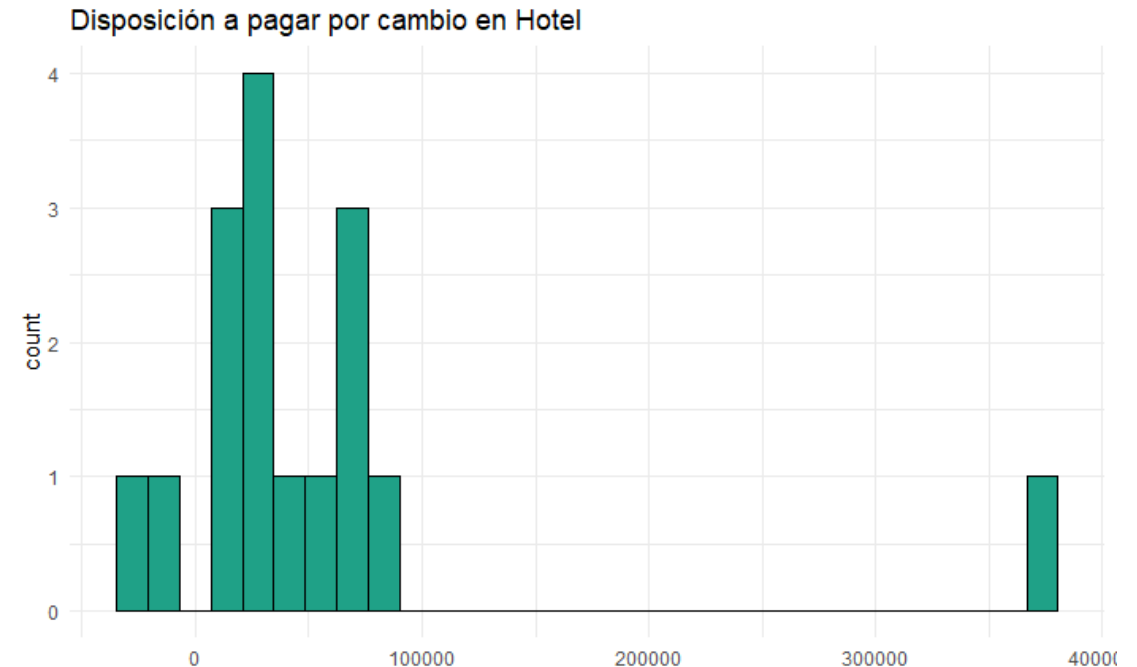
$$\text{Rango\_Precio} = \$120.000 - \$60.000 = \$60.000$$

$$\text{Rango\_uPrecio} = u(\text{Precio\_60k}) - u(\text{Precio\_120k})$$

Siendo el *willingness to pay*:

$$\text{Rango\_uHotel} * (\text{Rango\_precio} / \text{Rango\_uPrecio})$$

Puede observarse en este gráfico que la gran mayoría de los consumidores se encontrarían dispuestos a pagar entre \$5.000 y \$90.000 menos con respecto a la elección del hotel. Con algunos que se encontrarían dispuestos a pagar hasta \$20.000 pesos más al preferir un hotel de lujo. Cabe destacar que se excluyeron algunos outliers con un muy alto valor en el gráfico, que serán analizados en la pregunta 5.

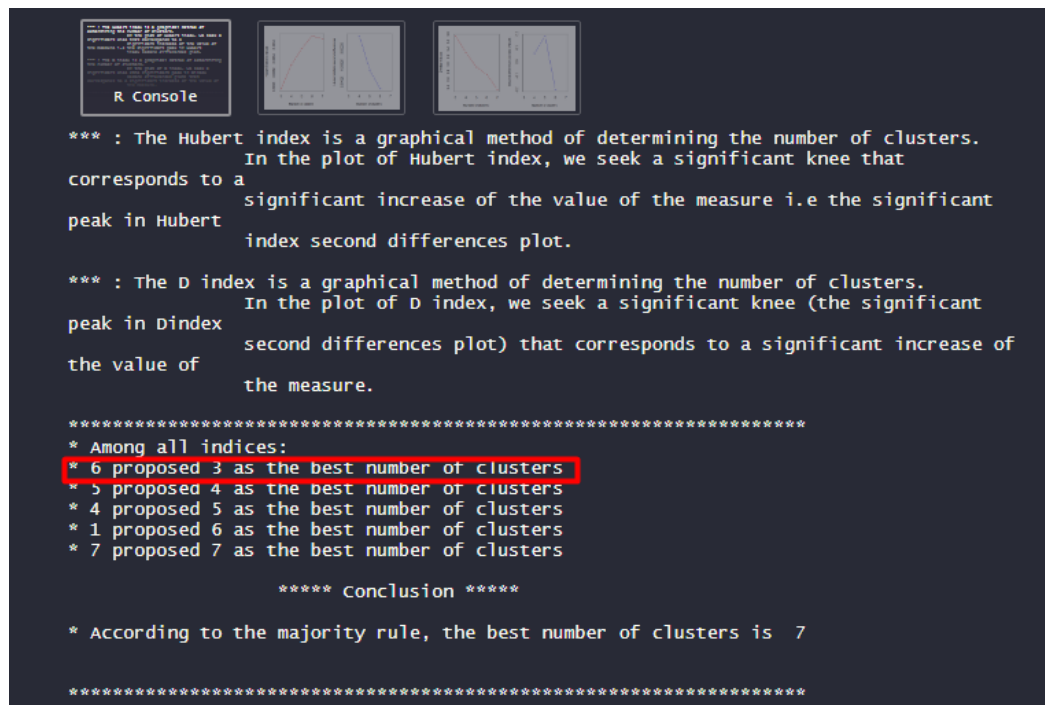




## Análisis de Datos Obtenidos

**5. Realice una segmentación de sus encuestados basándose en las preferencias obtenidas mediante la regresión. Comente respecto a los segmentos obtenidos. ¿Qué producto/servicio podría ofrecer a cada segmento?**

Para la segmentación de clientes se utilizó un método de aprendizaje no supervisado, haciendo uso del algoritmo K-Means con 3 clusters. Inicialmente se realizó un análisis para determinar el número óptimo de clusters utilizando la librería NbClust. Esta recomendó utilizar 7 clusters, sin embargo, se decidió utilizar 3 al ser la segunda mejor opción, además de que facilitará el análisis de una población tan reducida como esta.



```
R Console

*** : The Hubert index is a graphical method of determining the number of clusters.
      In the plot of Hubert index, we seek a significant knee that
corresponds to a significant increase of the value of the measure i.e the significant
peak in Hubert index second differences plot.

*** : The D index is a graphical method of determining the number of clusters.
      In the plot of D index, we seek a significant knee (the significant
peak in Dindex second differences plot) that corresponds to a significant increase of
the value of the measure.

*****
* Among all indices:
* 6 proposed 3 as the best number of clusters
* 5 proposed 4 as the best number of clusters
* 4 proposed 5 as the best number of clusters
* 1 proposed 6 as the best number of clusters
* 7 proposed 7 as the best number of clusters

***** conclusion *****

* According to the majority rule, the best number of clusters is 7

*****
```

## Análisis de Datos Obtenidos

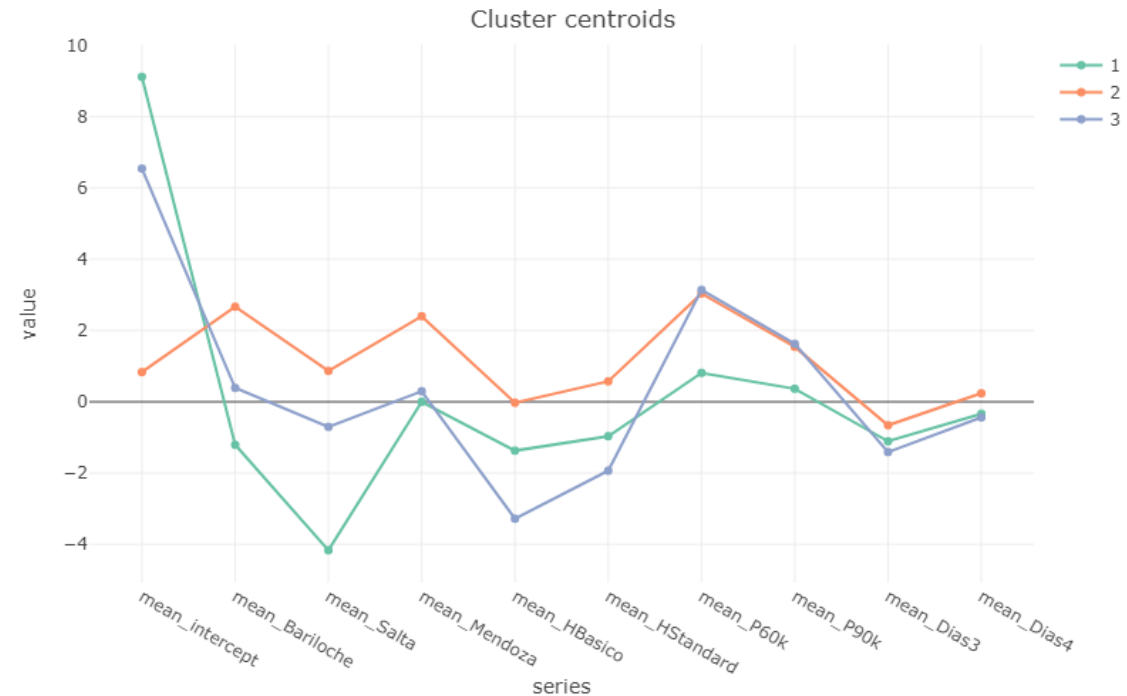
### 5. Realice una segmentación de sus encuestados basándose en las preferencias obtenidas mediante la regresión. Comente respecto a los segmentos obtenidos. ¿Qué producto/servicio podría ofrecer a cada segmento?

Al graficar los promedios para cada cluster, puede observarse que existen algunas diferencias entre ellos.

Al grupo número 1 se le podría ofrecer el paquete del intercepto con destino a El Chaltén, ya que se observa que las demás opciones representarían una menor utilidad para ellos, esta oferta podría venir acompañada de algún descuento por pertenecer a algún club exclusivo, ya que parece tratarse de un grupo con alto poder adquisitivo.

Puede observarse que Mendoza, Salta y Bariloche son los destinos favoritos para el grupo 2, a este grupo le puede resultar atractiva una oferta con estos destinos, especialmente si se combina con un hotel Standard y a un precio bajo, dando flexibilidad a elegir el número de días sin alterar demasiado el precio.

Al grupo 3 se les podría ofrecer un producto de precio bajo, siendo a Mendoza o a El Chaltén, este podría ser en un Hotel Standard o de Lujo, ajustando el número de días a 4.



## Análisis de Datos Obtenidos

**5. Realice una segmentación de sus encuestados basándose en las preferencias obtenidas mediante la regresión. Comente respecto a los segmentos obtenidos. ¿Qué producto/servicio podría ofrecer a cada segmento?**

Para complementar el análisis, también se realizó la segmentación utilizando el *willingness to pay* de cada grupo. Con lo que puede observarse que el grupo 1 ciertamente estaría bastante dispuesto a pagar mucho más por un hotel de lujo.

El grupo número 3 preferiría incluso ahorrarse hasta \$80.000 con respecto a un hotel de lujo, ya que, como vimos en el análisis anterior, un precio bajo representa una utilidad bastante alta para este grupo.

El 2 se encontraría dispuesto a pagar una suma pequeña, el producto propuesto para este grupo podría contener una opción de upgrade con un precio reducido para incluir un hotel de lujo.



## Análisis de Datos Obtenidos

**6. Genere dos perfiles de productos que no aparezcan en su cuestionario. Suponga que actualmente éstos son los únicos productos en el mercado ofrecidos por los competidores X e Y. Basándose en el total de sus encuestas, ¿cuál sería su estimación sobre la participación de mercado para cada uno de estos productos?**

Tras generar los nuevos productos, se realizó un cálculo para determinar la preferencia de los encuestados al respecto y así determinar su market share. Esto se realizó sumando los valores parciales de cada encuestado en relación a los atributos de los productos A y B. En caso de que la preferencia de A fuera mayor, se asignaba 1 a la columna “Perfil”, de lo contrario, se asignaba 0.

El análisis reveló una preferencia del 27% sobre el producto A y un 73% de preferencia para el producto B. Esto revela que la población encuestada en general tiende a preferir el Chaltén como destino turístico, por el mayor número de días posible.

En conclusión, El Chaltén es un destino turístico bastante atractivo, seguido por Mendoza y Bariloche. Esto indica que las 3 opciones mencionadas pueden representar una ventaja competitiva al momento de ofrecer este tipo de productos, combinándose con un precio asequible y diferentes opciones de hotel.

### Producto A

- **Hotel:** Standard
- **Precio:** \$90.000
- **Lugar:** Mendoza
- **Días:** 3

### Producto B

- **Hotel:** Básico
- **Precio:** \$60.000
- **Lugar:** El Chaltén
- **Días:** 5

ID	Intercept	Bariloche	Salta	Mendoza	El Chaltén	H_Basico	H_Standard	H_DeLujo	Precio_60k	Precio_90k	Precio_120k	Dias_3	Dias_4	Dias_5	WTP	Perfil
1	7.40540541	0.16666666666666668...	-1.9999999999999999...	0.6666666666666666...	0	-2.7297297	-1.9459459	0	3.00000000	1.2972973	0	-1.7837838	-1.05405405	0	54594.59	0
2	5.02702703	0.3333333333333333...	-0.6666666666666666...	2.0000000000000000...	0	-0.8108108	-0.5675676	0	3.45945946	1.6756757	0	-2.1081081	-1.72972973	0	14062.50	0
3	6.36036036	0.0000000000000000...	-0.6666666666666666...	0.0000000000000000...	0	-1.7837838	-0.5945946	0	3.21621622	2.8108108	0	-1.2432432	-0.48648649	0	33277.31	0
4	8.30630631	0.9999999999999999...	-0.6666666666666666...	-1.0000000000000000...	0	-7.4864865	-4.1351351	0	1.21621622	-0.1621622	0	-0.4054054	0.05405405	0	36933.33	1
5	0.72972973	2.6666666666666666...	0.0000000000000000...	1.6666666666666666...	0	-0.8378378	0.6216216	0	1.91891892	3.5135135	0	-0.1891892	0.78378378	0	26197.18	1
6	7.99099099	-0.5000000000000000...	-1.3333333333333333...	-3.0000000000000000...	0	-4.1621622	-2.3783784	0	3.40540541	1.9459459	0	-1.1351351	-0.64864865	0	73333.33	0
7	10.10810811	-1.6666666666666666...	-3.6666666666666666...	-0.9999999999999999...	0	-1.9189189	-1.5945946	0	-0.08108108	0.3243243	0	-1.0540541	-1.00000000	0	-142000.00	0
8	6.06306306	0.4999999999999999...	0.0000000000000000...	1.3333333333333333...	0	-1.7567568	-0.4594595	0	3.18918919	0.9189189	0	-3.5945946	-0.48648649	0	33050.85	0
9	7.78378378	0.8333333333333333...	-4.6666666666666666...	0.0000000000000000...	0	-0.6756757	-1.2972973	0	2.78378378	1.7837838	0	-1.3243243	-0.62162162	0	14563.11	0
10	7.32432432	-2.4999999999999999...	-3.3333333333333333...	1.3333333333333333...	0	-0.6486486	0.1081081	0	-1.21621622	1.5135135	0	0.2972973	0.97297297	0	-32000.00	1

## ANEXOS

### Utilidades Parciales por Usuarios



Utilidades Parciales por Usuario.zip

### Utilidades de Precio por Usuarios



Utilidades Precios Usuarios.zip

### Importancias Relativas por Usuarios



Importancias Relativas.zip