

Asignatura	Datos del alumno	Fecha
Análisis e Interpretación de Datos	Apellidos: Mondragón Guadarrama	31 – diciembre - 2020
	Nombre: José Carlos	

Actividades

Trabajo: Practicando con tablas de frecuencias y gráficos estadísticos

1. Observa la siguiente tabla y justifica porqué los valores que figuran en la fila «SUMA» son los mismos que los de la fila k de las frecuencias acumuladas absolutas y relativas.

Modalidades	Frecuencias	Frec. relativas	Frec. absolutas	Frec. relativas absolutas
1	n_1	f_1	N_1	F_1
2	n_2	f_2	N_2	F_2
...
k	n_k	f_k	N	1
SUMA	N	1		

Respuesta: las frecuencias acumuladas absolutas por su definición son la sumatoria de cada una de las frecuencias específicas cada dato hasta la modalidad (clase) que se necesite (Triola, 39, falta año). De esta forma, si se requiere conocer la frecuencia hasta la tercer case, hay que sumar las frecuencias del dato 1, las frecuencias del dato 2 y las frecuencias del dato 3.

$$Frec. Acumulada Absoluta = N_k = \sum_{i=1}^k n_i$$

k = total de las clases de los datos

La fila SUMA en la tabla, en la columna de frecuencias absolutas, debe mostrar la acumulación de cada una de las clases, incluyendo las anteriores. La primera frecuencia se puede interpretar como:

$$N_1 = \sum_{i=1}^{k=1} n_i = n_1$$

Siendo igual al primer renglón de la tabla. Para el segundo valor, el mismo procedimiento generará la siguiente sumatoria

$$N_2 = \sum_{i=1}^{k=2} n_i = n_1 + n_2$$

Asignatura	Datos del alumno	Fecha
Análisis e Interpretación de Datos	Apellidos: Mondragón Guadarrama	31 – diciembre - 2020
	Nombre: José Carlos	

Debería ser claro que para la clase T-1 donde T es el número total clases la ecuación debería interpretarse como:

$$N_{T-1} = \sum_{i=1}^{k=T-1} n_i = n_1 + n_2 + \cdots + n_{T-2} + n_{T-1}$$

Para el último valor se toma la suma N_{T-1} y se agrega la frecuencia de $k = T$.

$$N_T = \sum_{i=1}^{k=T-1} n_i + n_T = n_1 + n_2 + \cdots + n_{T-2} + n_{T-1} + n_T = \sum_{i=1}^{k=T} n_i$$

El hecho de ir sumando cada uno de los valores de la frecuencia es igual a tener cada valor de frecuencia de cada clase por separado y sumar al final todos, justo un caso específico para el total de las clases en el procedimiento de la columna de frecuencias en la fila suma. En dicha columna, una vez teniendo los datos hasta la clase k, se hace el resumen total sumando el n_1 mas el n_2 hasta llegar a n_k . De esta forma, la manera de obtener la suma de la columna es:

$$\text{Suma Frecuencias} = \sum_{i=1}^{k=T} n_i = n_1 + n_2 + \cdots + n_{T-1} + n_T$$

Por su parte, para la frecuencias relativas y frecuencias relativas absolutas, sucede un fenómeno parecido. Una frecuencia relativa es aquella en donde se obtiene la frecuencia de una clase y se divide entre la suma de todas las frecuencias(en el caso anterior N). Cada una de las filas en la columna de frecuencias relativas ha tomado el valor de frecuencias de su clase y la ha dividido entre N, de la siguiente forma:

$$f_i = \frac{1}{N} n_i$$

Caben destacar algunas propiedades:

- El valor de f_i nunca será mayor a 1 dado que en el caso donde solo existe una clase, $n_i = N$ y la división será el número entre él mismo.
- En el caso contrario, donde existen una infinidad de clases, N tenderá a infinito y f_i será un valor muy pequeño considerando una frecuencia cualquiera que sea entre un número excesivamente grande.

Asignatura	Datos del alumno	Fecha
Análisis e Interpretación de Datos	Apellidos: Mondragón Guadarrama	31 – diciembre - 2020
	Nombre: José Carlos	

La columna de suma utiliza la agregación para generar un resumen de las frecuencias relativas, indicando que se tienen el total computado. La fórmula para la sumatoria será:

$$\text{Suma Frecuencias Relativas} = \frac{1}{N}n_1 + \frac{1}{N}n_2 + \cdots + \frac{1}{N}n_{T-1} + \frac{1}{N}n_T = \frac{1}{N} \sum_{i=1}^{k=T} n_i$$

Por su parte, en la columna de frecuencias absolutas relativas, existe la interpretación de que el valor k (de la clase k) es la sumatoria de frecuencias relativas hasta esa clase (Tema 1: Introducción a la Estadística, maestría UNIR). Su fórmula es

$$F_k = \frac{1}{N} \sum_{i=1}^k n_i$$

Análogo a las frecuencias absolutas, el cálculo para la primer frecuencia relativa absoluta es:

$$F_1 = \frac{1}{N} \sum_{i=1}^{k=1} n_i = \frac{1}{N}n_1$$

De hecho este valor, será igual al primer valor de la columna de frecuencias relativas.

Para el segundo valor (k=2):

$$F_2 = \frac{1}{N} \sum_{i=1}^{k=2} n_i = \frac{1}{N}n_1 + \frac{1}{N}n_2$$

Para el penúltimo valor (T el total de clases):

$$F_{T-1} = \frac{1}{N} \sum_{i=1}^{k=T-1} n_i = \frac{1}{N}n_1 + \frac{1}{N}n_2 + \cdots + \frac{1}{N}n_{T-2} + \frac{1}{N}n_{T-1}$$

Caben destacar que dado que cada uno de los F_i nunca será mayor a 1 y el número representado como n_i forma parte del N total, la proporción de los datos será repartida en la ponderación de n_i .

Para el último valor:

$$F_T = \frac{1}{N} \sum_{i=1}^{k=T} n_i = \frac{1}{N}n_1 + \frac{1}{N}n_2 + \cdots + \frac{1}{N}n_{T-1} + \frac{1}{N}n_T$$

De esta manera debería ser claro que la suma de frecuencias relativas es igual al último valor de frecuencias relativas absolutas:

Asignatura	Datos del alumno	Fecha
Análisis e Interpretación de Datos	Apellidos: Mondragón Guadarrama	31 – diciembre - 2020
	Nombre: José Carlos	

$$\text{Suma Frecuencias Relativas} = \frac{1}{N} \sum_{i=1}^{k=T} n_i = F_T \frac{1}{N} \sum_{i=1}^{k=T} n_i$$

2. Fíjate en el ejemplo 4 donde figuran dos diagramas de barras. Ambos gráficos parecen idénticos pero no lo son. ¿Sabrías decir cuál es la diferencia?

Son dos visualizaciones de los datos. La gran diferencia es que el primero muestra las frecuencias absolutas y el segundo muestra las frecuencias relativas, es decir, su proporción con respecto al 100% de los datos. De cualquier manera es adecuado que ambos se parezcan y eso significa que el generador de la gráfica mantuvo la proporción de las escalas.

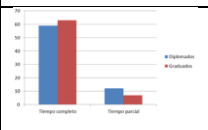
Los elementos principales para poder determinar la diferencia son:

- Los valores que cada columna ha incluido como etiqueta (ubicada cercana al tope de cada barra)
- La escala/valores del eje Y en la gráfica

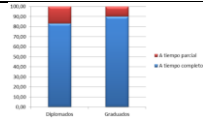
3. Partiendo del ejercicio anterior, ¿serías capaz de reconstruir la tabla de frecuencias a partir de los gráficos?

K	Frecuencias	Frecuencias relativas	Frecuencias absolutas	Frecuencias relativas absolutas
1 (Tiempo completo)	111	$111/127 = 0.874$	111	0.874
2 (Tiempo parcial)	16	$16/127 = 0.126$	127	1.0
Suma	127	1		

4. Observando los dos gráficos de barras del ejemplo 5: ¿Cuál de los dos gráficos anteriores te parece más adecuado? ¿Qué diferencias encuentras entre ellos? ¿Qué «pros» y contras le ves al uso de cada uno de ellos?

Gráfico	Características	Ventajas	Desventajas
	Clases = 2 x 2 {Diplomados, Graduados} X {Tiempo Completo, Tiempo Parcial} Medida base: frecuencia relativa	<ul style="list-style-type: none"> Muestra la <u>proporción</u> de graduados y diplomados en cada categoría 	No es claro si existen más graduados que diplomados.

Asignatura	Datos del alumno	Fecha
Análisis e Interpretación de Datos	Apellidos: Mondragón Guadarrama	31 – diciembre - 2020
	Nombre: José Carlos	

		(tiempo parcial y tiempo completo)	
		<ul style="list-style-type: none"> Es detallista 	
	<p>Clases = 2 x 2</p> <p>{Diplomados, Graduados} X {Tiempo Completo, Tiempo Parcial}</p> <p>Medida base: frecuencia absoluta</p>	<ul style="list-style-type: none"> Muestra claramente de un total, que proporción existe de cada una de las 4 posibilidades Recuenta claramente el 100% de los datos 	<p>La base particularmente para la segunda clase en cada columna está defasada lo que pudiera llegar a generar confusión.</p>

5. Una importante multinacional del sector tecnológico ha realizado pruebas de cociente intelectual entre sus empleados dentro de un proyecto de investigación de personal que están llevando a cabo. Deciden coger una muestra de 60 de ellos:

146	139	126	122	125	103	96	110	118	118
101	142	134	124	112	109	134	113	81	113
123	94	100	136	109	131	117	110	127	124
106	124	115	133	116	102	127	117	109	137
117	90	103	114	139	101	122	105	97	89
103	108	110	128	114	112	114	102	82	101

Como empleado de recursos humanos te piden realizar un informe a ti. ¿Sabrías graficar correctamente esta información previa elaboración de una tabla adecuada? ¿A la hora de hacer la tabla, cuantas posibilidades tienes de organizarla? Argumenta porqué empleas un tipo de grafica esta información de cara al informe y no otro.

Respuesta: Los datos anteriores no significan nada si no son colocados en contexto. Un primer acercamiento es determinar cuáles son los rangos de IQ interesantes para alguien que quiera “evaluar” a los empleados. Para esto, se utilizó la escala WAIS-III (Wechsler, 1997) por ser una de las más referidas. Cabe destacar que sin el dato de la escala que se utilizó esto no es más que un mero ejercicio y se aleja de un posible reporte real.

Asignatura	Datos del alumno	Fecha
Análisis e Interpretación de Datos	Apellidos: Mondragón Guadarrama	31 – diciembre - 2020
	Nombre: José Carlos	

Los datos crudos a priori, deberán generar una tendencia hacia el número 116.5, que es el promedio (utilizando N-1 datos, dado que son menos de 100), y teniendo la mayor parte de los datos entre 102 y 132 (de acuerdo a la desviación estándar. Para ello, se puede utilizar una gráfica de dispersión



Así se puede analizar cuál es la media de los empleados y 1 desviación estándar hacia adelante así como una hacia atrás. Aquellos empleados con bajo IQ, en este caso la muestra 19 y 59, son elementos con los que hay que trabajar fuertemente. Sin embargo, cabe la posibilidad de, a pesar del bajo score, en realidad sean personas que pueden desempeñar cabalmente su tarea por que ni siquiera se conoce si esto es alto o bajo, solo se establece que están por fuera de una cantidad de datos.

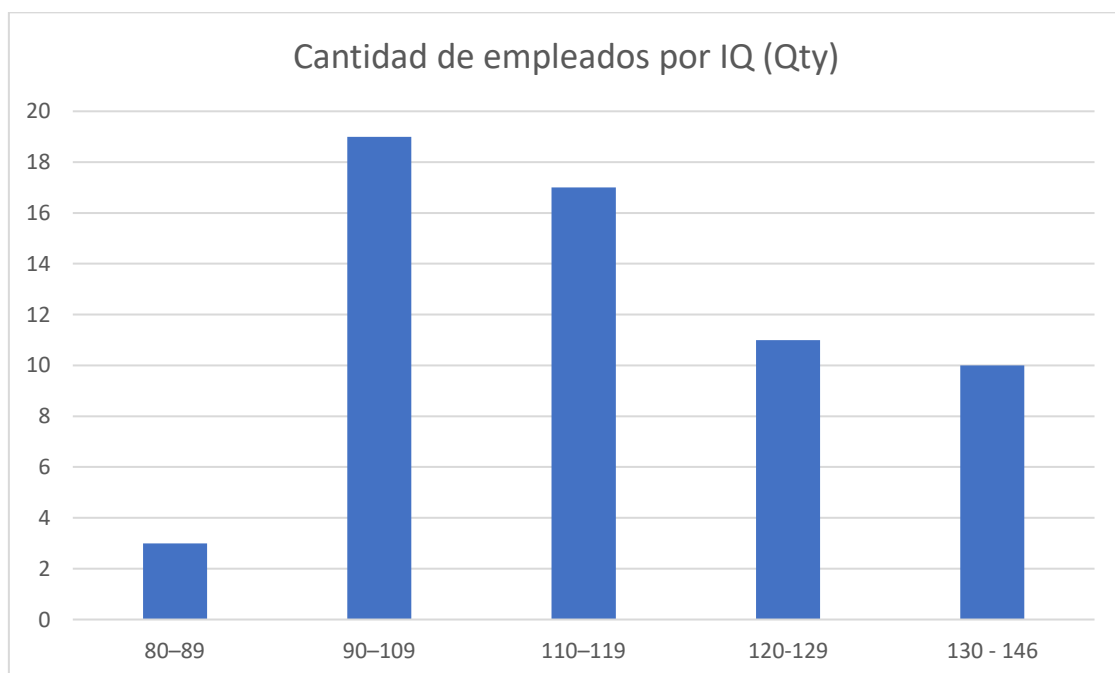
Para colocar en contexto, se agregó la calificación de WAIS-III, con lo que la información muestra:

Clase		Frecuencia	Frq. Relativa	Frq. Abs	Freq.Abs Rel
	Lim inf	Lim sup			
69 and below	40	69	0	0	0
70–79	70	79	0	0	0
80–89	80	89	3	0.05	0.05
90–109	90	109	19	0.31666667	0.36666667
110–119	110	119	17	0.28333333	0.65
120-129	120	129	11	0.18333333	0.83333333
130 - 146	130	146	10	0.16666667	1
		Suma=	60	1	

Asignatura	Datos del alumno	Fecha
Análisis e Interpretación de Datos	Apellidos: Mondragón Guadarrama	31 – diciembre - 2020
	Nombre: José Carlos	

1. De acuerdo a la escala, 3 empleados están por debajo del promedio. Particularmente los dos previamente discutidos están bastante cerca de una clasificación que el autor llama “Borderline”.
2. Hay una buena cantidad de empleados entre 90 y 119, esperado por ser el promedio y un poco encima del promedio.
3. Finalmente existen alrededor de 34% de empleados que están muy por encima del promedio (el tope se estableció en 146, no por la clasificación sino porque fue el mayor dato que se encontró en la muestra).

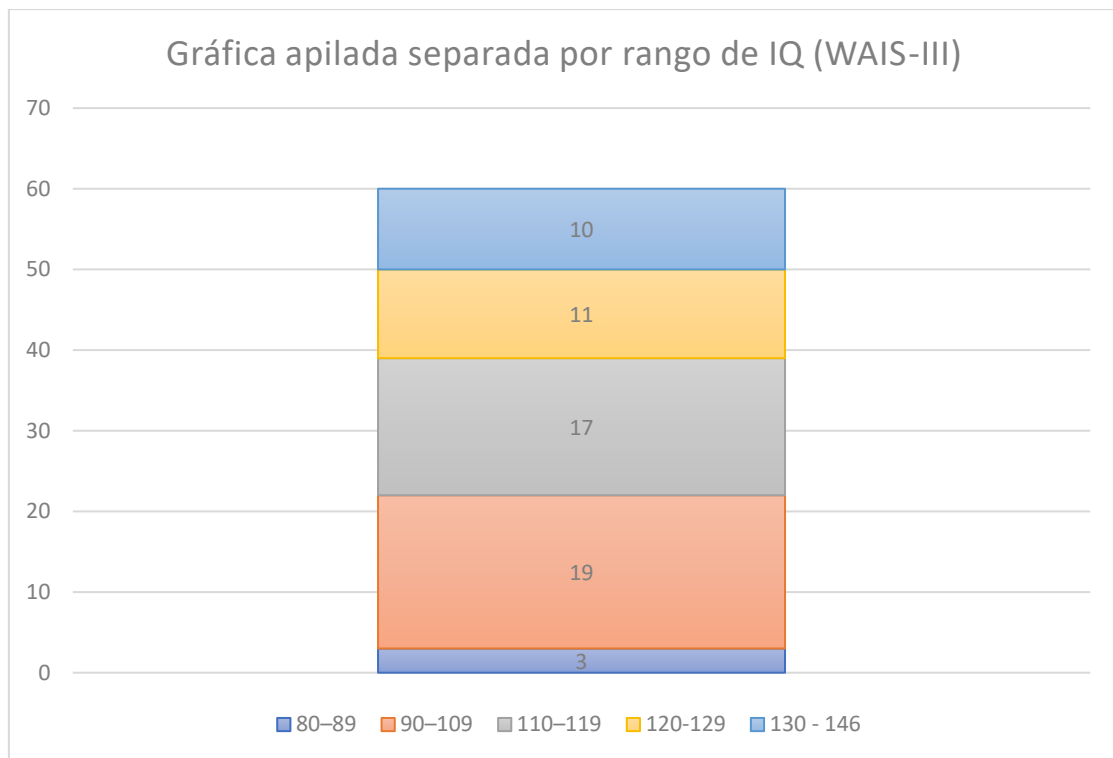
Para mostrar mejor esta tabla se decidió una gráfica de barras dado que las clases al final ya se establecen como una variable cualitativa (aunque ordinal) y será directamente visible la proporción que se tienen de un segmento de empleados contra otro.



(gráfica de cantidad de empleados por IQ, asimétrica positiva)

Así mismo, también es interesante ver el 100% de los empleados, y sus diferencias (absolutas, no proporcionales) contra el resto de los grupos. Para poder analizar esta situación se utilizó un gráfico de barras apilado que claramente muestra que el grupo 2 y 3 (clases 2 y 3, promedio y arriba de promedio) son la mayoría.

Asignatura	Datos del alumno	Fecha
Análisis e Interpretación de Datos	Apellidos: Mondragón Guadarrama	31 – diciembre - 2020
	Nombre: José Carlos	

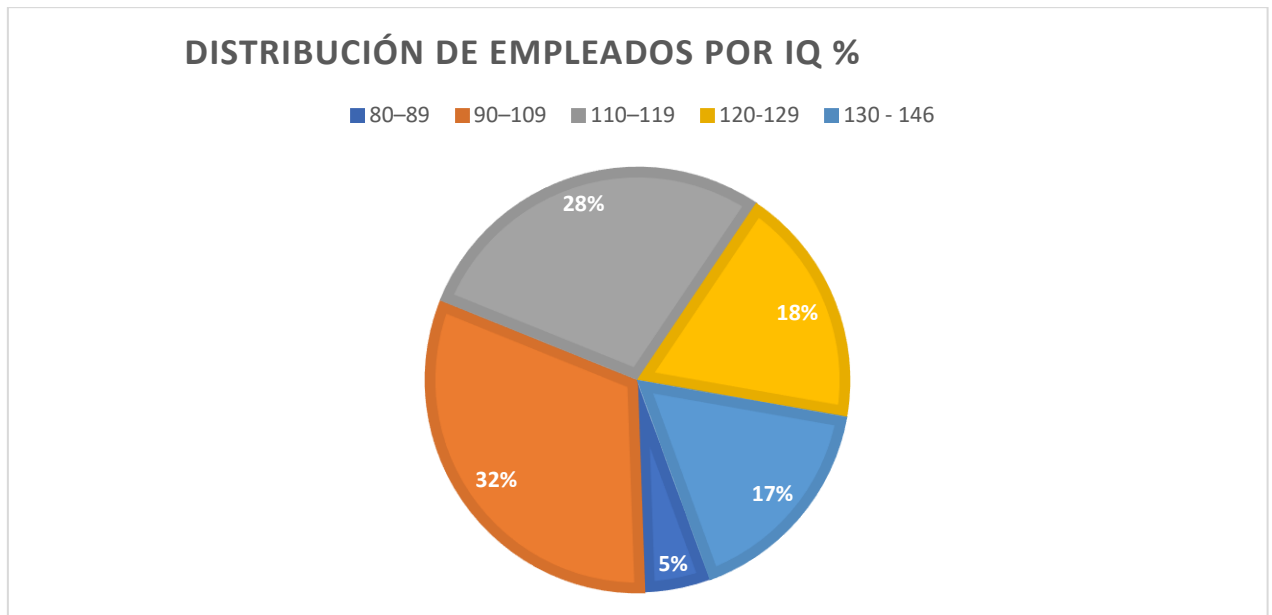


6. Partiendo del ejercicio anterior: has realizado el informe donde has empleado dicha gráfica, pero un compañero tuyo de la empresa al verla te pregunta: «¿por qué no empleaste un gráfico de sectores para representarla?» ¿Qué le responderías? Razona tu respuesta. ¿Piensas que es imposible emplear un gráfico de sectores para representar dicha información? ¿Se te ocurre algún cambio que pudieras hacer en la organización de los datos para que tal cosa fuera posible?

La sugerencia es adecuada puesto que la relación proporcional de los grupos ayuda a determinar acciones para las mayorías, si es que hubiera alguna. El analizar los porcentajes podría indicar situaciones críticas en particular con proporciones altas en rangos inferiores. Sin embargo, esconde la cantidad de personas que realmente cuentan con un cierto nivel.

Para generar la gráfica, basta con utilizar las frecuencias relativas vistas en la tabla anterior y aplicar el algoritmo de construcción.

Asignatura	Datos del alumno	Fecha
Análisis e Interpretación de Datos	Apellidos: Mondragón Guadarrama	31 – diciembre - 2020
	Nombre: José Carlos	



7. Revisa el vídeo del apartado «Lecciones magistrales» para realizar la última actividad. En resumen, esta actividad consistirá en que confecciones tu propio informe *Analytics* a partir de una emulación basada en un auténtico informe de Google Analytics (ten en cuenta que se trata de una práctica algo más extensa que las actividades anteriores).

Se tomó como base el registro de visitas obtenido de:

https://micampus.unir.net/files/1398314/download?download_frd=1

Objetivo: describir comportamientos de los usuarios y del tráfico teniendo como ejes base la hora de la visita (y a partir de la hora, el día), el id. De usuario y la clasificación de la visita como rebote y/o nueva.

Características de cada visita:

Campo	Descripción
Hora	minutos desde el inicio o puesta en funcionamiento del sitio.
Duración visita	Minutos que estuvo el usuario en el sitio
Usuario	Id. Único del usuario
Páginas Vistas	Cantidad de páginas (referencias http distintas) que visitó
Rebote	Indicador que clasifica 1 si el usuario salió (rebote) o se quedó en el sitio (0).
Visita nueva	Indicar que clasifica 1 si la visita no fue identificada antes.

Asignatura	Datos del alumno	Fecha
Análisis e Interpretación de Datos	Apellidos: Mondragón Guadarrama	31 – diciembre - 2020
	Nombre: José Carlos	

Columnas agregadas

Con los datos anteriores, se pueden obtener las siguientes columnas derivadas para un mejor análisis:

Campo	Descripción
Día	<p>Día de la visita basado en 0. El primer día es el día 0 y así en la serie. Se tuvieron un total de 32 días.</p> <p>Fórmula para obtenerlo:</p> $dia = \left\lfloor \frac{[horadevisita/60]}{24} \right\rfloor$ <p>Es importante revisar que las operaciones están truncadas/limitadas por el operador de la función piso. De esta forma se discretiza tanto la hora como el día en el que se generó la visita.</p>
Hora Base 24	<p>Hora del día en el que se dio la visita. La primer hora es la 0 y sigue hasta la hora 23.</p> <p>Fórmula para obtenerlo:</p> $dia = [horadevisita/60] \bmod 24$ <p>El operador mod entendido como el residuo de dividir a/b. mod = residuo(a,b)</p> <p>De esta forma la hora de visita lineal desde el primer minuto del sitio en línea se traduce a días de 24 horas.</p>

Análisis primario de los datos

A partir de los elementos anteriores, los datos muestran el siguiente comportamiento:

	Día	Hora de visitas	Hora base 24	Duración visita	Usuario	Páginas vistas	Rebote	Visita Nueva
s	9.308	13397.93	7.103	15.388	139.266	2.019	0.353	0.491
μ	15.916	23624.23	11.237	15.062	257.108	3.068	0.146	0.408
Σ				10845		2209	105	294

s = desviación estándar

μ = media poblacional

Σ = sumatoria de los datos

No se muestran las columnas de suma para aquellas columnas que son derivadas o que son parte de una serie de tiempo que refiere el momento en el que se dio el evento

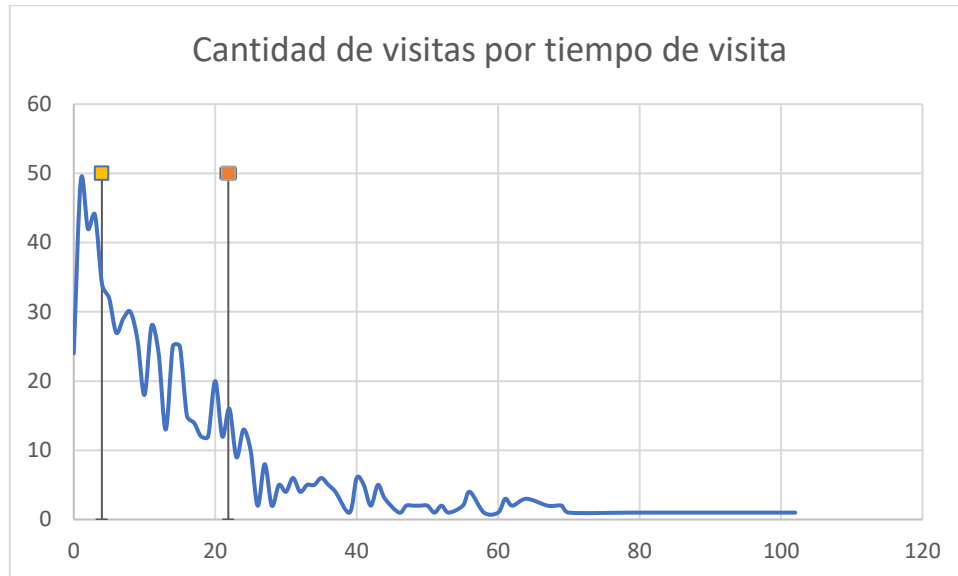
Asignatura	Datos del alumno	Fecha
Análisis e Interpretación de Datos	Apellidos: Mondragón Guadarrama	31 – diciembre - 2020
	Nombre: José Carlos	

1. La hora de visitas, el Usuario y el Día no se tomará en cuenta para el análisis primario
2. La hora promedio de visita al sitio es a las 11 am, sin embargo los datos se encuentran esparcidos con una desviación estándar de 7.1. Esto significa que las visitas se encuentran a lo largo de todo el día en este conjunto de datos.
3. La duración promedio de la visita es de 15 minutos, sin embargo, la desviación estándar es bastante grande con lo que una visita típicamente podría estar entre los 0 y los 30 minutos.
4. Los usuarios que acceden al sitio en promedio visualizan 3 páginas. Es importante destacar que en los datos, los usuarios de rebote no son contabilizados, sin embargo, para que este valor sea adecuado, hay que contabilizar como población solo las 615 entradas que no son rebote, variando un poco la desviación estándar. Si bien el promedio cambia de 3.068 a 3.592 la desviación estándar se acorta a 1.699.
5. La media del análisis del rebote para datos que son 0 y 1 corresponde con la proporción de usuarios que sí son rebote. En este sentido hay un 14.6% de usuarios que son considerados rebote y coincide con la media (1's entre el total de la población)
6. Igual que el rebote, se puede indicar que un 40.8% de las visitas son usuarios nuevos. Hay que considerar que de este porcentaje hay un 6.2% de visitas que fueron rebote. De esta manera un número más adecuado sería que existen 34.6% de visitas orgánicas reales nuevas.

Asignatura	Datos del alumno	Fecha
Análisis e Interpretación de Datos	Apellidos: Mondragón Guadarrama	31 – diciembre - 2020
	Nombre: José Carlos	

Análisis de cuartiles

1. Cuartiles para tiempo de visita en el sitio



Gráfica x: tiempo en el sitio y: frecuencia de tiempo en el sitio

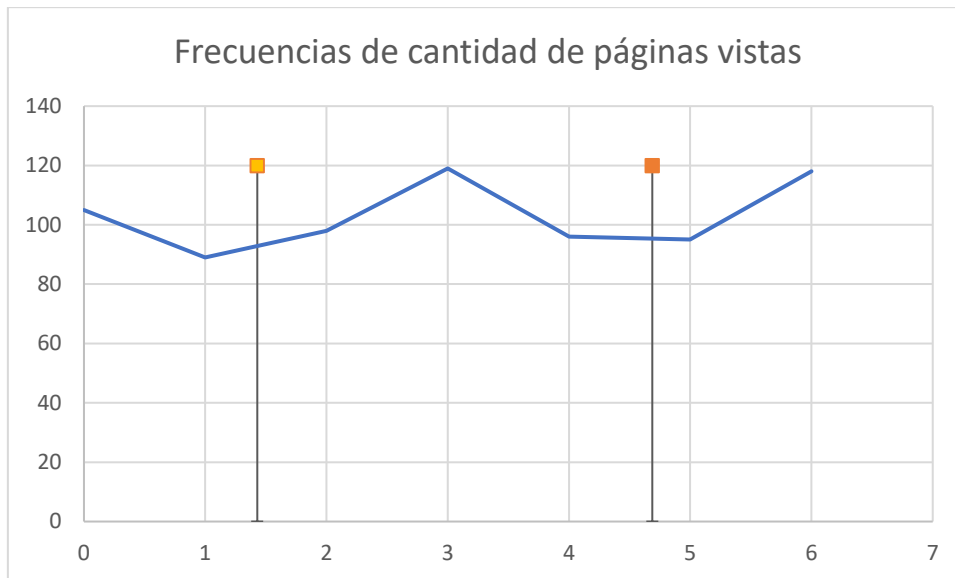
$Q1 = 3.945$, $Q2 = 10.55$, $Q3 = 21.83$

De esta forma, 25% de los usuarios normalmente pasan menos de 4 minutos en el sitio, 50% de los usuarios pasan poco mas de 10 min y el 75% de los usuarios pasaran 21.83 min. En la gráfica solo han sido marcados Q1 y Q3.

Además la gráfica claramente muestra una asimetría positiva asegurada por el cálculo del As de Pearson de: 0.914. Cabe destacar lo anterior para no esperar que los usuarios se queden en el sitio una cantidad de tiempo mayor a 10 minutos. Para comprobar esto basta con medir la diferencia entre el valor de Q1 y Q2 (6.605) y Q2 a Q3 (11.28). El cociente de estos números para medir su relación es de 1.7, es decir, tuvieron que pasar 1.7 veces la diferencia de Q1 y Q2 para alcanzar el 75% de los datos.

Asignatura	Datos del alumno	Fecha
Análisis e Interpretación de Datos	Apellidos: Mondragón Guadarrama	31 – diciembre - 2020
	Nombre: José Carlos	

2. Cuartiles para páginas visualizadas en el sitio



$Q_1 = 1.428$, $Q_2 = 3.387$, $Q_3 = 4.687$

De esta forma, 25% de los usuarios normalmente ven menos de 2 páginas en el sitio, 50% de los usuarios ven menos de 4 y el 75% de los usuarios no llegarán a 5 páginas vistas. En la gráfica solo han sido marcados Q_1 y Q_3 .

Además la gráfica claramente muestra leve asimetría positiva asegurada por el cálculo del As de Pearson de: 0.033.

Preguntas importantes alrededor de los datos

Las siguientes son algunos cuestionamientos que pueden ser determinantes para cálculos de presupuestos, inversiones o incluso viabilidad de proyectos relacionados con el sitio:

Pregunta	Justificación
¿Cómo se reparten las visitas en el día, tanto usuarios recurrentes como únicos, por separado y en conjunto? Y ¿Qué hora del día tiene más visitas?	En el caso de hostings como el de Amazon que cuentan con tarifas dinámicas, el saber la distribución de visitas en el día ayudará a optimizar la cantidad de recursos que se usen para atenderlos.
¿Cuánto contenido consumen usuarios no nuevos?	Para los usuarios que han estado regresando, es importante ofrecer contenido de retención. Esto genera ingresos a partir de Google ads o cualquier otra estrategia de monetización.
¿Para usuarios que entraron 1 sola vez, cuánto contenido consumen?	Dadas las nuevas visitas, es importante que pasen de la página base. Si lo hacen entonces el contenido es interesante.

Asignatura	Datos del alumno	Fecha
Análisis e Interpretación de Datos	Apellidos: Mondragón Guadarrama	31 – diciembre - 2020
	Nombre: José Carlos	

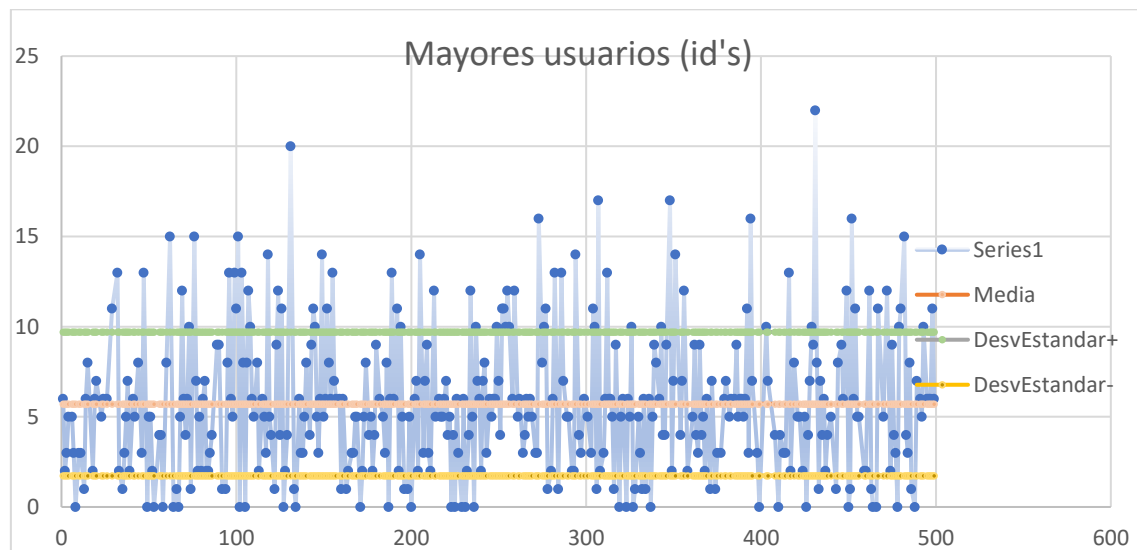
¿Qué cantidad de minutos de visualización existe por día y por hora?	La modificación del contenido y su manera de retener a los usuarios son claves para la monetización del sitio.
¿Qué cantidad de páginas son visualizadas por día y por hora?	La modificación del contenido y su manera de retener a los usuarios son claves para la monetización del sitio.
¿Qué usuarios son los mayores consumidores?	Teniendo identificados a los “clientes” del sitio, es posible la activación de campañas de “premios” para aquellos fieles que muestren al resto de las personas las ventajas de estar dentro del producto.

Asignatura	Datos del alumno	Fecha
Análisis e Interpretación de Datos	Apellidos: Mondragón Guadarrama	31 – diciembre - 2020
	Nombre: José Carlos	

Análisis de usuarios en el sitio

Revisando el sitio con más detenimiento, podemos responder las preguntas orientadas a los usuarios.

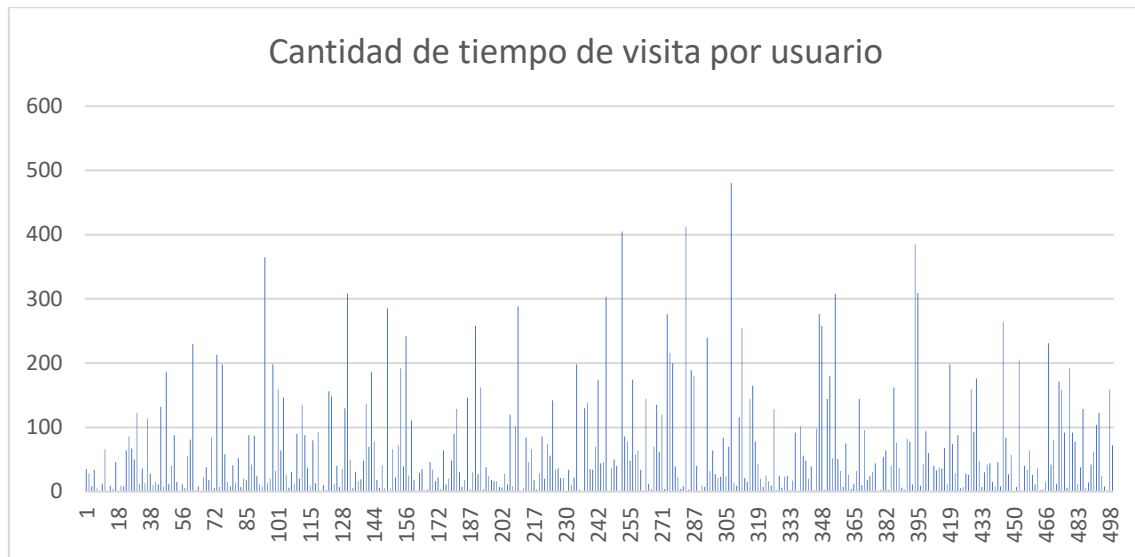
La siguiente gráfica muestra las frecuencias de cada usuario, ordenado por ID de usuario y mostrando la cantidad de página vistas por. Cabe destacar que existen dos usuarios que han entrado una mayor cantidad, y que son claramente representados en esta gráfica. Hay una variación importante en los datos y el objetivo del sitio a largo plazo debería estar orientado hacia subir a todos los usuarios por debajo de la media, para desplazar los 2 estadísticos hacia arriba (media y desviación estándar).



Sumando todo el tiempo que ha entrado cada usuario, se puede identificar durante este periodo de tiempo, cuánta es la inversión en minutos de cada uno. Esta gráfica toma en cuenta usuarios que no fueron rebotados (por ende es una gráfica que acerca más a la visita orgánica real).

De cualquier forma, la media de estos datos está en 64.7, indicando que los usuarios que ya están enganchados, visualizan una gran cantidad de contenido (basta con analizar la diferencia entre 5 de los mayores usuarios y el promedio), sin embargo, la gran mayoría no pasará en un mes, más de una hora de visualización. Lo anterior se puede constatar identificando una desviación estándar de 78.5, un muy amplio rango.

Asignatura	Datos del alumno	Fecha
Análisis e Interpretación de Datos	Apellidos: Mondragón Guadarrama	31 – diciembre - 2020
	Nombre: José Carlos	



Ya dentro de los usuarios que han visitado el sitio más de una vez, los datos muestran un comportamiento diferente. La media para estos datos está en 102.98 mientras que su desviación estándar llega a 86. Es decir, para este tipo de usuarios, aunque su promedio de tiempo en el sitio es mayor, uno podría esperar que la desviación estándar fuera un poco menor, pero genera más variación. Dado que es difícil analizarlo, se tomó el coeficiente de variación (desv. Estándar dividido por la media y multiplicado por 100%)

Coeficiente de variación de los usuarios en general: 121.33%

Coeficiente de variación de los usuarios en usuarios recurrentes: 83.51%

De esta manera es más claro que para usuarios con más de una visita, su experiencia ya no varía tan fuerte como en la de usuarios en general.



Es interesante analizar que dada la hora del día en la que el usuario entra, existe una relación con la cantidad de tiempo que pasará en el sitio. Aunque es difícil saber el porqué

Asignatura	Datos del alumno	Fecha
Análisis e Interpretación de Datos	Apellidos: Mondragón Guadarrama	31 – diciembre - 2020
	Nombre: José Carlos	

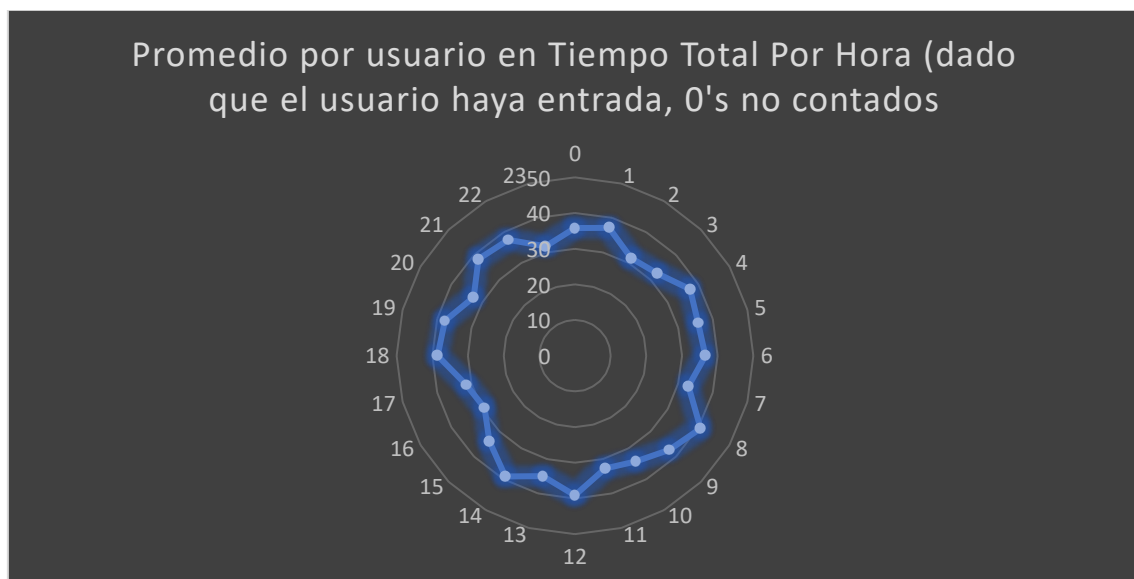
de esta situación, mecanismos de publicidad pueden estar en función de esta distribución y maximizar su objetivo. Los datos se obtuvieron identificando de todos los usuarios NO REBOTE, cuánto era el mínimo del tiempo que pasaban en cada día. Posteriormente, se obtuvo el mínimo para todos los usuarios y se presentó la información.

De aquí es claro que los usuarios, al entrar cerca de la hora 17, hay una alta probabilidad de que pasen al menos 5 minutos en el sitio.



Finalmente, para el sitio es importante saber el tiempo que un usuario, al haber entrado el sitio va a estar dentro de él. Para esto se muestra con respecto a la hora del día, la cantidad de tiempo promedio de los usuarios (SIN CONTAR REBOTES).

De esta manera se analiza que la mayor cantidad de tiempo promedio está en las horas 8 y 14. Así, los usuarios que entren aproximadamente en esos tiempos, pasarán alrededor de 40 minutos visualizando información.



Asignatura	Datos del alumno	Fecha
Análisis e Interpretación de Datos	Apellidos: Mondragón Guadarrama	31 – diciembre - 2020
	Nombre: José Carlos	

Análisis de visualización en el sitio por fecha y hora

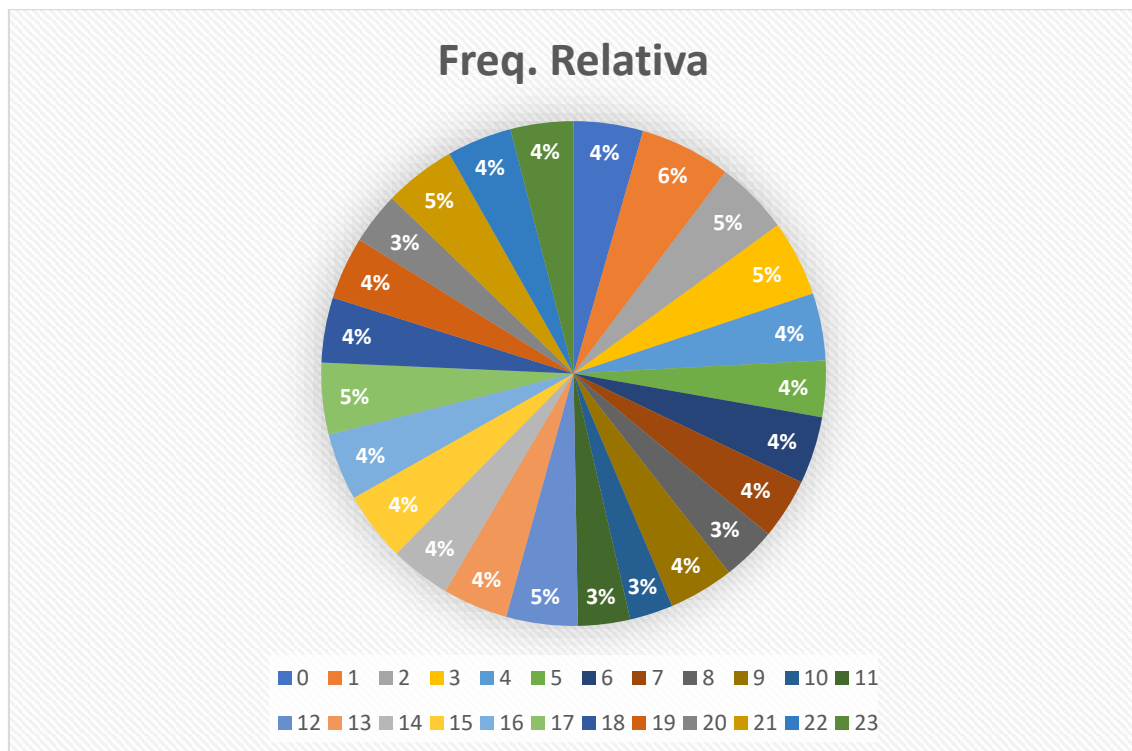
Por su parte, y cambiando la perspectiva de análisis, una manera diferente de identificar oportunidades en las visitas, es determinando la distribución que existe a cada hora. Aunque en la sección anterior se introdujeron algunas de estas métricas, enseguida se mostrarán a más profundidad.

La cantidad de usuarios en el día, separados por hora, se muestra en la siguiente grafica. Es muy claro que a la hora 1, este dato es mayor llegando a ser entre 40 y 45 visitas. La media está en 30 visitas con una desviación estándar de 4, de manera que la mayor cantidad de horas, las visitas estarán entre 26 y 34, teniendo una carga baja para servidores.



La cantidad de visitar que se tienen a esa hora es mayor completando un 6% de la carga total del sitio. Esto no indica que la cantidad de horas que los usuarios pasan durante ese tiempo es alta (para no contraponerse con las medidas anteriores), solo que es una frecuencia mayor.

Asignatura	Datos del alumno	Fecha
Análisis e Interpretación de Datos	Apellidos: Mondragón Guadarrama	31 – diciembre - 2020
	Nombre: José Carlos	



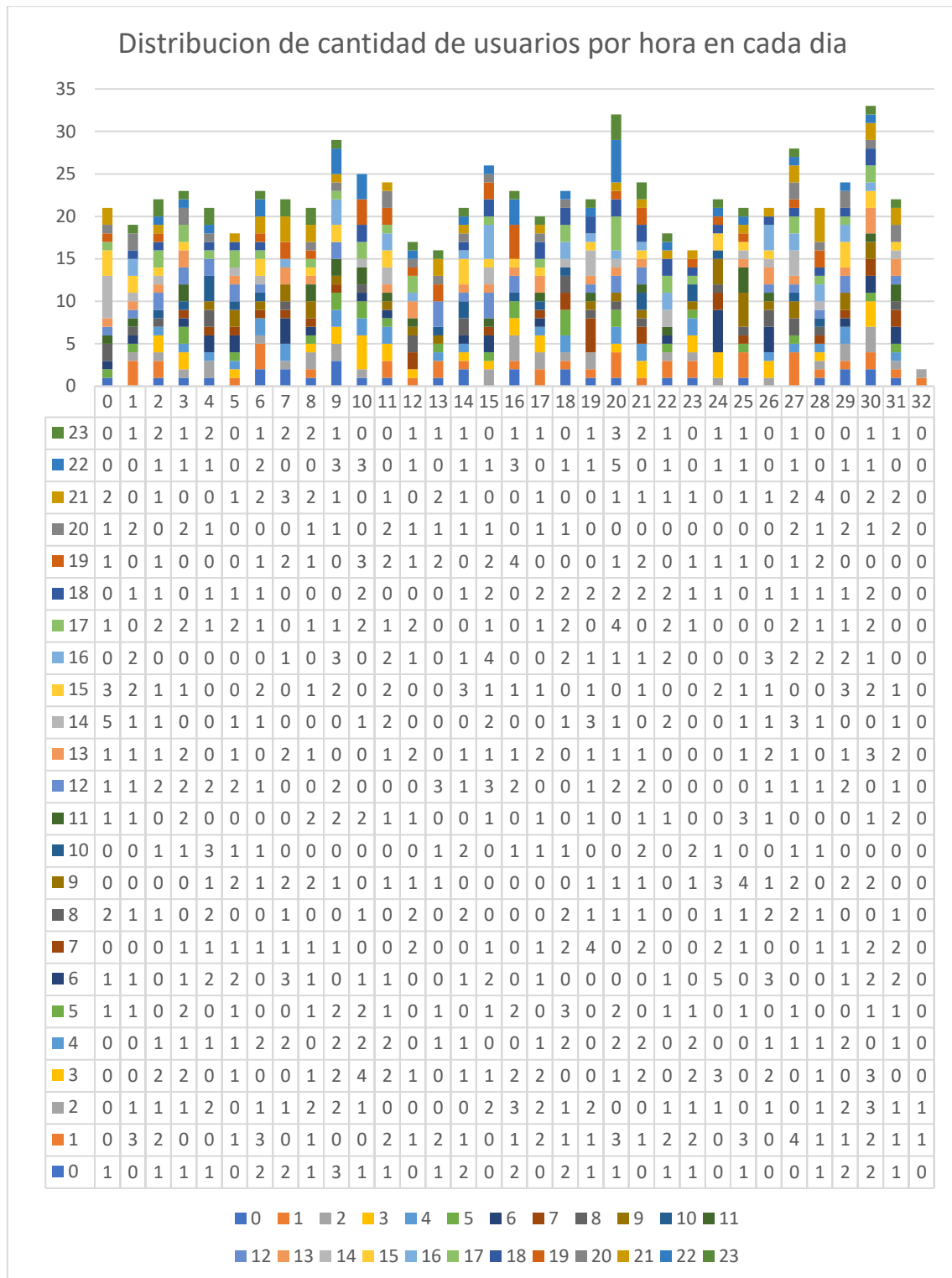
Identificar los días fuertes de carga en el sitio pueden dar efectos sobre campañas generadas, o sobre tendencias alrededor de los tiempos de vida de las personas. En esta gráfica es claro que dentro de los primeros 8 días (semana 1) la variación entre la cantidad de usuarios entrando es poca, llegando por debajo de 20 y hasta 23. Después de la primer semana que se podría tomar como un primer ciclo, las diferencias se hacen mayores, con dos días que incluso llegan a bajos históricos para ese momento. Sin embargo, los picos y el promedio durante este periodo es mayor por hasta 1 usuario (21 usuarios vs 22 usuarios). Podría parecer poco 1 usuario, pero multiplicado por la cantidad de días, ya son diferencias relevantes. Al día 30 se encuentra un máximo histórico de 33 que para la última semana es congruente. Se puede identificar que durante los últimos 6 días (sin contar el 32 cuyos datos son incompletos), el promedio de usuarios ronda los 24, con mínimos por encima de 20.

Asignatura	Datos del alumno	Fecha
Análisis e Interpretación de Datos	Apellidos: Mondragón Guadarrama	31 – diciembre - 2020
	Nombre: José Carlos	



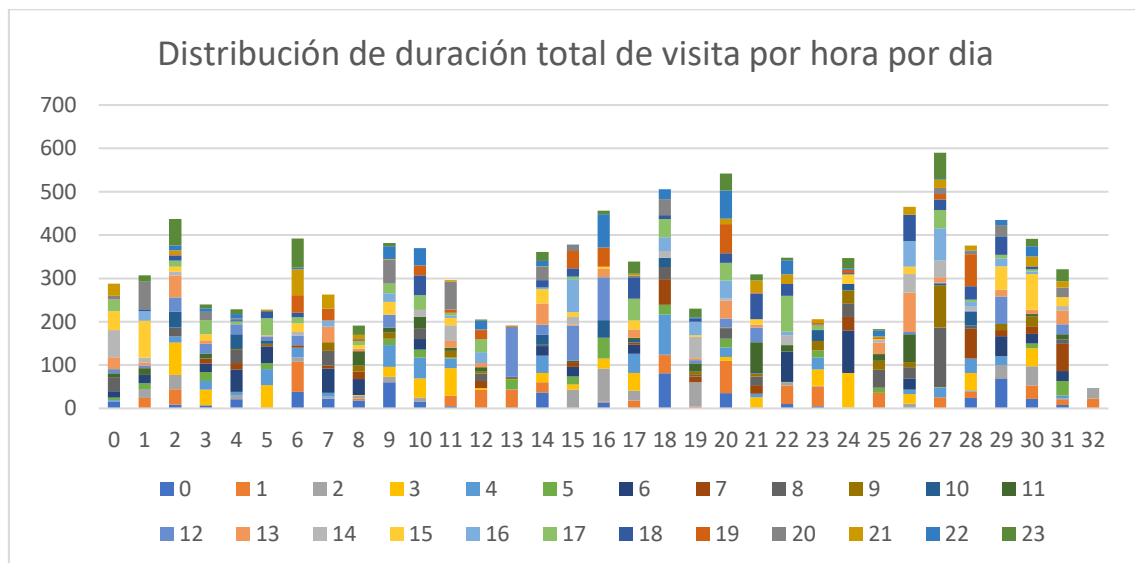
Recapitulando la información anterior, la siguiente gráfica muestra la distribución de usuarios en el día por cada hora. La gráfica de barras apilada permite identificar simultáneamente la cantidad de personas totales en ese día, y la hora a la que llegaron más. Es importante destacar que esto es por cantidad de usuarios, no determina ni el ancho de banda utilizado ni la cantidad de páginas o el tiempo que pasaron. Un día necesario a repetir en el futuro es el 30 en donde en promedio tuvo más de un usuario por hora.

Asignatura	Datos del alumno	Fecha
Análisis e Interpretación de Datos	Apellidos: Mondragón Guadarrama	31 – diciembre - 2020
	Nombre: José Carlos	



Asignatura	Datos del alumno	Fecha
Análisis e Interpretación de Datos	Apellidos: Mondragón Guadarrama	31 – diciembre - 2020
	Nombre: José Carlos	

Finalmente, y tratando de sumar información a la gráfica anterior, se puede analizar la cantidad total de tiempo que pasaron los usuarios en el sitio, y la hora a la que se registró su primer ingreso. Si bien la gráfica anterior habla de captación, ésta habla de retención. Es interesante ver que el día 30 no fue el de mayor retención, sino el 27. En la gráfica anterior es también un día muy aceptable en el rango solo 5 usuarios por detrás del líder pero con más de 100 minutos de diferencia en el tiempo de visita.



Nota para resolver los ejercicios: Puede que no haya solo una posibilidad de respuesta correcta, lo interesante es argumentar porque se piensa así en algunos casos.

Bibliografía

Wechsler. (1997). *WAIS-III: Wechsler adult intelligence scale*. San Antonio TX: Psychological Corporation.