UNIVERSIDAD DE LOS ANDES

TAREA 3

Almacenamiento y Procesamiento Masivo de Datos

Juan de Dios Valenzuela 15/11/2017

Esta tarea consistirá en procesar los reviews de Yelp, y generar un archivo CSV o similar con datos que permita analizar y comparar, con un notebook Jupyter, la distribución de las estrellas recibidas por todos los restaurantes en cada categoría.

Contenido

Resultados	2
Gráficos	2
Resultados Obtenidos	11
Proceso	12
Conclusiones:	14
Fuentes y referencias	15
Anexos:	16

Resultados

Gráficos

Preschools http://public.lightning-viz.org/visualizations/ae6d0c0b-83e1-4fcd-b6b1-eecb30619a69/public/

(Anexo 1)

Apartments http://public.lightning-viz.org/visualizations/674e8ddd-f5f7-4a56-9e60-be216fba01d1/public/

(Anexo 2)

Asian Fusion http://public.lightning-viz.org/visualizations/ff3cc341-adab-4fca-96c2-24cbc8b23920/public/

(Anexo 3)

Auto Repair http://public.lightning-viz.org/visualizations/13fe9cf7-6cb6-4306-85e7-829dfe3ccede/public/

(Anexo 4)

American (Traditional) http://public.lightning-viz.org/visualizations/4475bdbf-20d9-468b-8311-202ac463fcc8/public/ (Anexo 5)

Arts & Entertainment http://public.lightning-viz.org/visualizations/b2f960fc-0783-4250-9a4a-87904e085d79/public/

(Anexo 6)

Automotive http://public.lightning-viz.org/visualizations/e7d762fe-538f-4098-91e3-4d0bd9105786/public/

(Anexo 7)

Bakeries http://public.lightning-viz.org/visualizations/c14b48e9-e1a3-4605-b93c-884e0490e69c/public/

(Anexo 8)

Barbeque http://public.lightning-viz.org/visualizations/0dfbb8cf-cc34-4539-a22b-5bc753eb8333/public/

(Anexo 9)

Bars http://public.lightning-viz.org/visualizations/860d4033-5838-4133-aa26-0f7dab83d35f/public/

(Anexo 10)

Beauty & Spas http://public.lightning-viz.org/visualizations/058d5f1f-4626-4d49-926d-a0dba3a733f7/public/

(Anexo 11)

Brasseries http://public.lightning-viz.org/visualizations/14403a9c-6de5-4edb-add8-a479eacced89/public/ (Anexo 12)

Breakfast & Brunch http://public.lightning-viz.org/visualizations/0af002b1-5ac3-4022-80ef-f5e9cc16aee2/public/ (Anexo 13)

Breweries http://public.lightning-viz.org/visualizations/25477726-3ade-4d0e-b705-00f0c75cad38/public/

(Anexo 14)

Burgers http://public.lightning-viz.org/visualizations/f8ebe047-f042-410c-9adc-184365c405d3/public/

(Anexo 15)

Cafes http://public.lightning-viz.org/visualizations/5e4a0a8a-7922-4e73-a7bd-b52ab430d275/public/

(Anexo 16)

Car Dealers http://public.lightning-viz.org/visualizations/7d2746a5-659e-40bc-9461-e533919902d2/public/

(Anexo 17)

Caribbean http://public.lightning-viz.org/visualizations/65972115-6f03-448d-be01-654897a62e66/public/

(Anexo 18)

Child Care & Day Care http://public.lightning-viz.org/visualizations/13ece310-8a25-4569-8a72-a366c19667ef/public/

(Anexo 19)

Chinese http://public.lightning-viz.org/visualizations/d0bcc3ff-8663-4614-a438-095962e98487/public/

(Anexo 20)

Cocktail Bars http://public.lightning-viz.org/visualizations/811b3392-4e73-4cbb-aebf-9371fc494a5d/public/

(Anexo 21)

Hair Salons http://public.lightning-viz.org/visualizations/21c6496f-5a32-4362-9af7-293a4c200b25/public/

(Anexo 22)

Coffee & Tea http://public.lightning-viz.org/visualizations/d0aff467-5e8a-4902-b54c-4c0603af8536/public/

(Anexo 23)

Contractors http://public.lightning-viz.org/visualizations/4c100710-5d52-4935-bcb3-6bd9061fa664/public/

(Anexo 24)

Cosmetic Dentists http://public.lightning-viz.org/visualizations/05583e94-6ad0-4cd8-890c-e8d16c9cd528/public/
(Anexo 25)

Delis http://public.lightning-viz.org/visualizations/d8224dd8-2f1a-4f37-81d0-421b5d073dca/public/

(Anexo 26)

Dentists http://public.lightning-viz.org/visualizations/b5a54cf1-cebe-432f-a8d0-5841513c9408/public/

(Anexo 27)

Diners http://public.lightning-viz.org/visualizations/58a4f2f2-17a9-4703-a06c-8b939c640422/public/

(Anexo 28)

Drugstores http://public.lightning-viz.org/visualizations/942e3993-3275-4a21-a7c9-b5bbcdcc728a/public/

(Anexo 29)

Education http://public.lightning-viz.org/visualizations/5c36f5c6-fca8-459c-83e9-9f328a21027f/public/

(Anexo 30)

Educational Services http://public.lightning-viz.org/visualizations/a2c25483-e7e7-4016-ad10-6069218c68d3/public/

(Anexo 31)

Electronics http://public.lightning-viz.org/visualizations/b6e14489-771d-4f6d-a5c4-fdac92534539/public/

(Anexo 32)

Elementary Schools http://public.lightning-viz.org/visualizations/b9c875ef-03f5-4644-9a72-50341667f721/public/

(Anexo 33)

Event Photography http://public.lightning-viz.org/visualizations/6fc6cbfe-87b2-418d-9a25-06cc40c8ced3/public/

(Anexo 34)

Event Planning & Services http://public.lightning-viz.org/visualizations/ed5f86b4-5d43-4a5b-8d0f-ba1826ea1b75/public/ (Anexo 35)

Fast Food http://public.lightning-viz.org/visualizations/0814492e-c4ed-4c12-92ec-9f6bce04d74a/public/

(Anexo 36)

Fish & Chips http://public.lightning-viz.org/visualizations/65632749-566b-4824-b39a-e773ed28940a/public/

(Anexo 37)

Food http://public.lightning-viz.org/visualizations/c6aaeb62-c603-4e9f-aa70-a0f1804ca190/public/

(Anexo 38)

French http://public.lightning-viz.org/visualizations/adbf89a5-0cc8-4cdc-af12-ecfe75c27a82/public/

(Anexo 39)

General Dentistry http://public.lightning-viz.org/visualizations/591b625b-d885-475e-8219-ee0966bfbe44/public/

(Anexo 40)

Gluten-Free http://public.lightning-viz.org/visualizations/2cc766ed-5593-4728-9fc9-88d966895060/public/

(Anexo 41)

Hair Stylists http://public.lightning-viz.org/visualizations/fc337673-e654-4dad-9f7c-0f32a382ac6c/public/

(Anexo 42)

Health & Medical http://public.lightning-viz.org/visualizations/e3282a8a-e030-4aa1-94d3-103ea1f75c94/public/

(Anexo 43)

Home Automation http://public.lightning-viz.org/visualizations/f4842609-6f02-4f43-b1ad-58ff4d82c9f3/public/

(Anexo 44)

Home Services http://public.lightning-viz.org/visualizations/d3b8f6ca-df74-4e0a-9c18-d0d5530cec27/public/
(Anexo 45)

Home Theatre Installation http://public.lightning-viz.org/visualizations/e2d677c6-91c2-433a-bc0e-e75756d4f592/public/ (Anexo 46)

Hotels & Travel http://public.lightning-viz.org/visualizations/69618b34-db3a-4ca8-94e5-6700f6425a4f/public/

(Anexo 47)

Hotels http://public.lightning-viz.org/visualizations/4f34132f-921a-4600-b379-5d97b2151738/public/

(Anexo 48)

Ice Cream & Frozen Yogurt http://public.lightning-viz.org/visualizations/d9364f22-66fe-45c6-9cd2-5791deb69415/public/

(Anexo 49)

Indian http://public.lightning-viz.org/visualizations/0259e17b-0110-48b6-b231-d38640b99175/public/

(Anexo 50)

Italian http://public.lightning-viz.org/visualizations/657a394a-e8d6-4930-9371-a45f645841c4/public/

Japanese http://public.lightning-viz.org/visualizations/6c7b3926-b89a-4ba8-829b-ed6561c47f2e/public/

(Anexo 51)

Language Schools http://public.lightning-viz.org/visualizations/95e024fe-8914-4c46-91bd-5d8383b083bb/public/ (Anexo 52)

Local Services http://public.lightning-viz.org/visualizations/c13f46e1-163c-421e-9431-19c66fc00969/public/ (Anexo 53)

Massage Therapy http://public.lightning-viz.org/visualizations/313f8b4a-27ce-4120-a972-87d3dcee689a/public/ (Anexo 54)

Massage http://public.lightning-viz.org/visualizations/f5fa09c5-da65-4a0b-b918-02da5a23a7ed/public/
(Anexo 55)

Meat Shops http://public.lightning-viz.org/visualizations/6ae2f470-9b44-462b-acf6-fb1f5ae98171/public/

(Anexo 56)

Mexican http://public.lightning-viz.org/visualizations/83b9d0d9-e5e5-4197-a227-dc2b4482c54b/public/

(Anexo 57)

Middle Schools & High Schools http://public.lightning-viz.org/visualizations/b10bdda6-04f6-48be-9204-828bcd7ea797/public/ (Anexo 58)

Modern European http://public.lightning-viz.org/visualizations/70632601-1b1c-477b-928f-6ad8355af557/public/

Nail Salons http://public.lightning-viz.org/visualizations/f5b2cbc6-ef0d-4d69-8b1d-1b19ebcef691/public/

(Anexo 59)

Specialty Food http://public.lightning-viz.org/visualizations/3cad1e70-bd8c-4b7b-927c-94759f41172d/public/

(Anexo 60)

Nightlife http://public.lightning-viz.org/visualizations/51234100-22b9-4f08-b532-4f19d1887191/public/

(Anexo 61)

Performing Arts http://public.lightning-viz.org/visualizations/8676a558-9057-4ab2-9376-0ad8a1ac1300/public/

(Anexo 62)

Periodontists http://public.lightning-viz.org/visualizations/3783b5b6-2c79-41e4-83ef-6af4190fdffd/public/

(Anexo 63)

Photographers http://public.lightning-viz.org/visualizations/20d7dfb5-3b44-45f9-af26-44c278e205fa/public/

(Anexo 64)

Pool Halls http://public.lightning-viz.org/visualizations/807f004c-b45b-4888-a351-e28206f0b45c/public/

(Anexo 65)

Preschools http://public.lightning-viz.org/visualizations/2aadb531-5c0e-4eb0-a917-4805b12c9366/public/

(Anexo 66)

Professional Services http://public.lightning-viz.org/visualizations/936f358e-bea6-4838-ae9d-9d2642eb9363/public/

(Anexo 67)

Real Estate http://public.lightning-viz.org/visualizations/3ea1ce3e-aeff-4856-b0ef-60e349337be8/public/

(Anexo 68)

Reiki http://public.lightning-viz.org/visualizations/732d0bdd-0fb9-4b4e-bf7f-bc377ab5adcc/public/

(Anexo 69)

Restaurants http://public.lightning-viz.org/visualizations/6d435c1a-391c-4fa1-b041-bc9e966c158b/public/

(Anexo 70)

Salad http://public.lightning-viz.org/visualizations/31f5b51b-5204-48ad-a03a-4e21feba7ff6/public/

(Anexo 71)

Sandwiches http://public.lightning-viz.org/visualizations/a56a0548-474e-4d13-ba49-0b97a197d512/public/

(Anexo 72)

Seafood http://public.lightning-viz.org/visualizations/f69a5631-1704-403e-9bff-9f282b94d981/public/

(Anexo 73)

Security Systems http://public.lightning-viz.org/visualizations/a5580c61-5a38-49c9-a546-7de23ce8b479/public/
(Anexo 74)

Session Photography http://public.lightning-viz.org/visualizations/94260632-58a9-4769-a09a-203662837219/public/

Shades & Blinds http://public.lightning-viz.org/visualizations/7b29591c-4d4b-48a8-9040-ef578a6b1c8a/public/

(Anexo 75)

Shopping Centers http://public.lightning-viz.org/visualizations/79134184-d75e-4966-aa53-ae37ec82985f/public/

(Anexo 76)

Shopping http://public.lightning-viz.org/visualizations/e08c6253-4c74-4e5f-8edd-7a5f2a0b07dc/public/

(Anexo 77)

Soul Food http://public.lightning-viz.org/visualizations/d220976d-b323-47bb-b37e-31b4cbb48103/public/

(Anexo 78)

Specialty Schools http://public.lightning-viz.org/visualizations/444d20eb-b124-442e-a7d2-833f51762b21/public/

(Anexo 79)

Sporting Goods http://public.lightning-viz.org/visualizations/11430aee-17ca-4000-8aed-fd620984263b/public/

(Anexo 80)

Steakhouses http://public.lightning-viz.org/visualizations/6bc050af-73bf-4a69-a826-78d40327ec21/public/

(Anexo 81)

Sushi Bars http://public.lightning-viz.org/visualizations/2685b2db-72b4-4d82-9d4e-511b07a834f4/public/

(Anexo 82)

Tanning http://public.lightning-viz.org/visualizations/4375b84f-defd-4f95-b622-aab30803aca3/public/

(Anexo 83)

Tapas Bars http://public.lightning-viz.org/visualizations/de7a1bea-5e8b-497a-9de0-95822db59714/public/

(Anexo 84)

Tours http://public.lightning-viz.org/visualizations/93a957e9-5816-4b66-ac8c-d2d9a2ba2868/public/

(Anexo 85)

TV Mounting http://public.lightning-viz.org/visualizations/0b3d921c-277c-4ef0-b477-fdfe97f05ae5/public/

(Anexo 86)

Vegetarian http://public.lightning-viz.org/visualizations/431497ea-3bb5-4f38-af6a-cb19d5360771/public/

(Anexo 87)

Japanese http://public.lightning-viz.org/visualizations/8d661660-4b4f-4776-a8df-0e488875521a/public/ (Anexo 88)

Nail Salons $\underline{\text{http://public.lightning-viz.org/visualizations/432bdadf-6b44-4913-9656-0649e19cb5c7/public/}$

(Anexo 89)

Shades & Blinds http://public.lightning-viz.org/visualizations/5fd58b2b-1fbf-443f-8885-4ba792a2babb/public/

(Anexo 90)

Resultados Obtenidos

Datos	Tiempo
5.000	0.483000040054 seconds
10.000	4.48699998856 seconds
30.000	IOError: [Errno 22] Invalid argument
Todo el Dataset	IOError: [Errno 22] Invalid argument

Proceso

Se comenzó descargando *Lightning* mediante *pip* de Python 2.7. Tras investigar un error sobre *ipython*, me di cuenta de que era una librería básica para realizar gráficos. (Librería que utiliza *pyplot*).

```
python -m pip install lightning-python
```

Posterior a esto, se instaló la API para funcionar como cliente:

```
python -m pip install jupyter
```

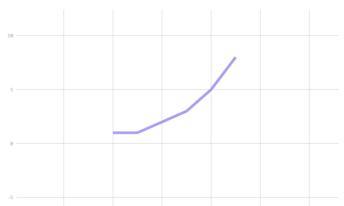
Teniendo todo instalado, se prosiguió a correr *jupyter* mediante el comando:

```
jupyter notebook
```

Para comprobar que estuviese todo bien instalado, se graficó un ejemplo de forma local, teniendo éxito al abrir el navegador con la dirección: http://public.lightning-viz.org/visualizations/91e26832-e664-413f-aad9-b5e0488485f9/public/, mostrando la recta programada.

```
from lightning import Lightning

lgn = Lightning(host='http://public.lightning-viz.org')
viz = lgn.line([1, 1, 2, 3, 5, 8])
viz.open()
```



Luego se prosiguió a estudiar los parámetros necesarios y comandos para la creación de un histograma.

Al completar los pasos anteriores, se escribió un programa en Python 2.7 que a partir del archivo *new_business.json* y *new_review.json* se generara un archivo .csv con las estrellas obtenidas por cada categoría.

Para este proceso, se escribió un programa de ayuda para generar los archivos new_business.json y new_review.json a partir de los archivos reales, el cual a partir de un parámetro n, se obtiene los primeros n business_id del archivo review.json, y luego de obtienen las k entradas del archivo business.json que tengan dichos id. (Se habla de k entradas ya que no todos los business_id del archivo review.json se encuentran en el archivo business.json). k < n

Además de este programa, se escribió otro con el fin de leer el archivo .cvs, y abrir un histograma por cada línea contenido en él.

Conclusiones:

¿Qué pasa con sus scripts cuando se incrementa el número de comentarios a revisar?

Se puede notar un incremento en el tiempo de ejecución. Esto debido a que debe procesar una mayor cantidad de datos, por lo que el tiempo entre procesos es bastante más caro. Se debe utilizar mayor cantidad de memoria, por lo que para los casos en que se requería procesar una cantidad de datos mayor a 10.000, este provocaba la caída del programa al llenar el buffer.

¿Detectó algún patrón interesante al revisar los datos visualmente?

Al no entregar la cantidad de *bins* como parámetro a la función del histograma, este tendría la cantidad de barras según la cantidad de veces en que haya existido al menos una persona que calificara con n estrellas. Esto provoco que algunos histogramas distribuyeran el eje x con valores con decimales. A medida que el número de comentario a revisar aumentaba, los gráficos tendían a tener 5 barras con el eje x con valores enteros, ya que existe al menos una persona que calificara con x estrellas a dicha categoría, para x entre 0 y 5.

Fuentes y referencias

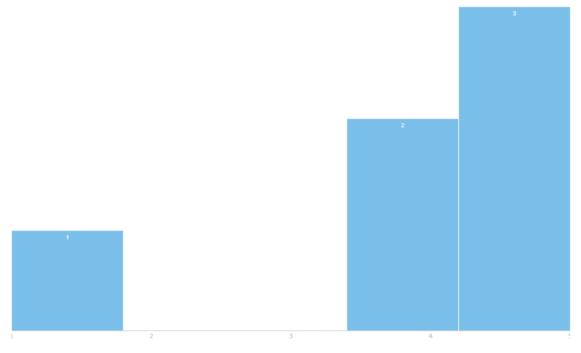
http://lightning-viz.org/usage/#creating

http://lightning-viz.org/usage/#client

http://nbviewer.jupyter.org/github/lightning-viz/lightning-example-notebooks/blob/master/misc/severless.ipynb

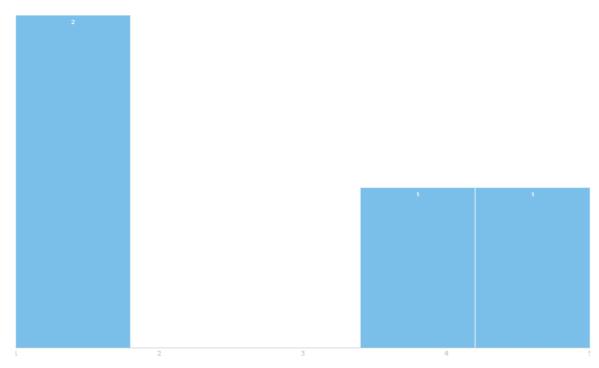
Anexos:

Anexo 1:



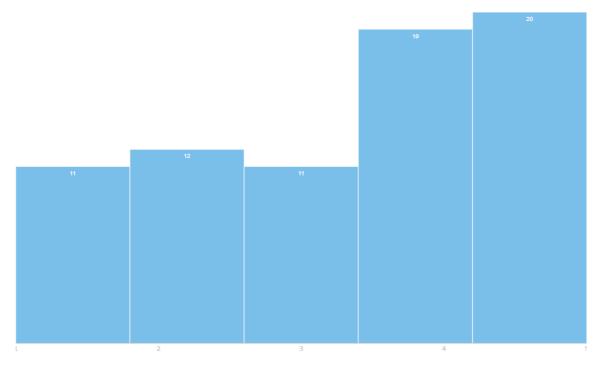
Categoria: Preschools

Anexo 2:



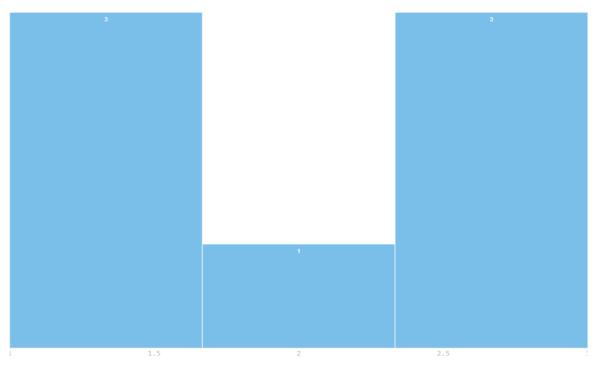
Categoria: Apartments

Anexo 3:



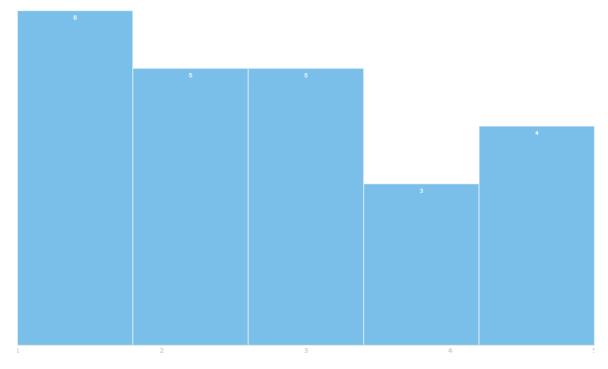
Categoria: Asian Fusion

Anexo 4:



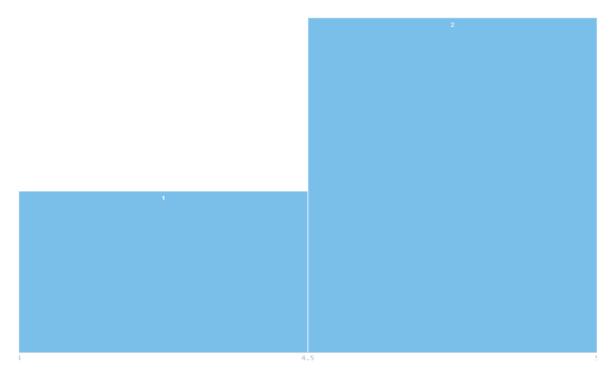
Categoria: Auto Repair

Anexo 5:



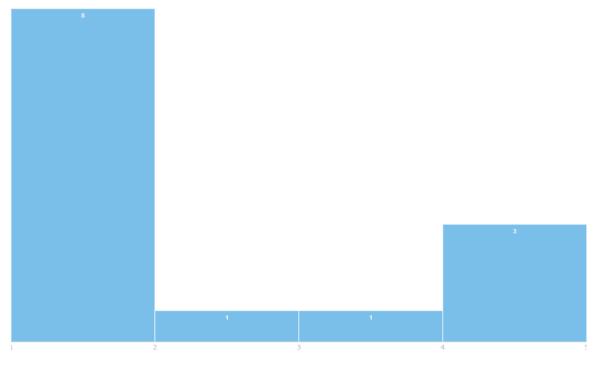
Categoria: American (Traditional)

Anexo 6:



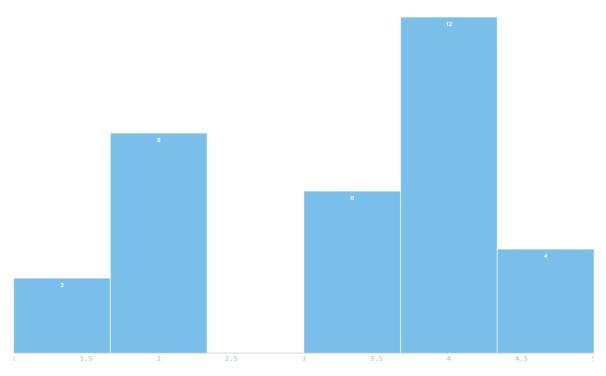
Categoria: Arts & Entertainment

Anexo 7:



Categoria: Automotive

Anexo 8:



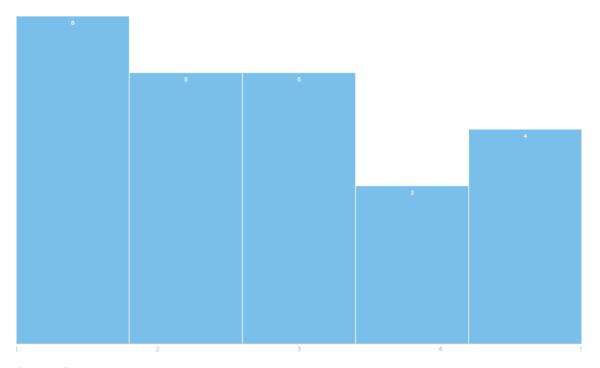
Categoria: Bakeries

Anexo 9:



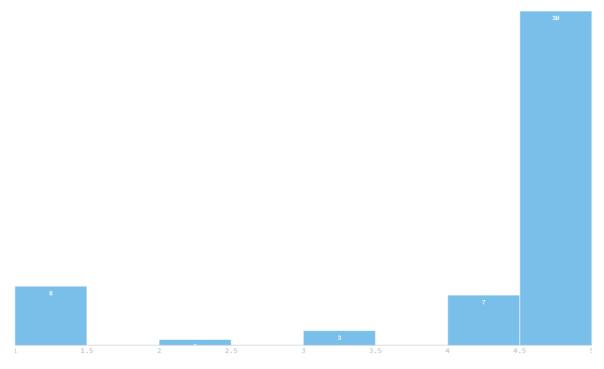
Categoria: Barbeque

Anexo 10:



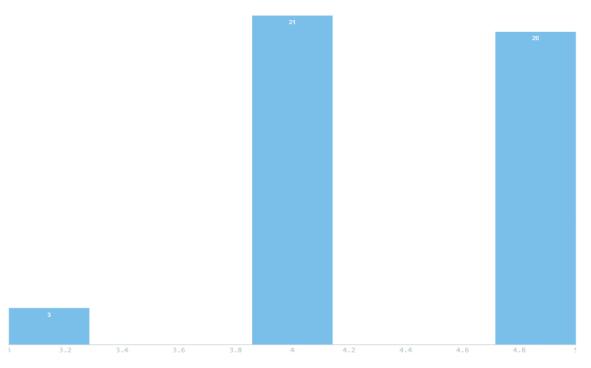
Categoria: Bars

Anexo 11:



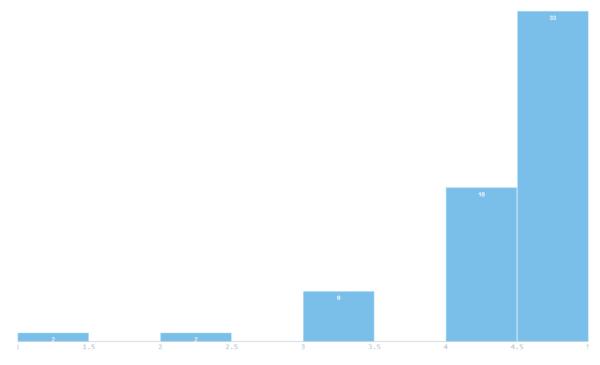
Categoria: Beauty & Spas

Anexo 12:



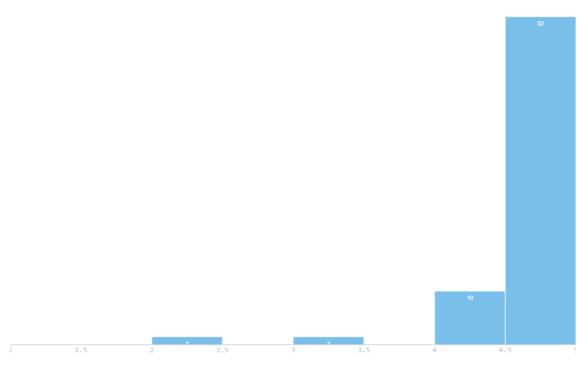
Categoria: Brasseries

Anexo 13:



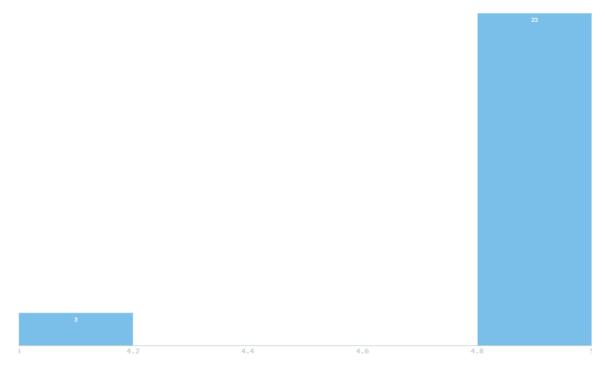
Categoria: Breakfast & Brunch

Anexo 14:



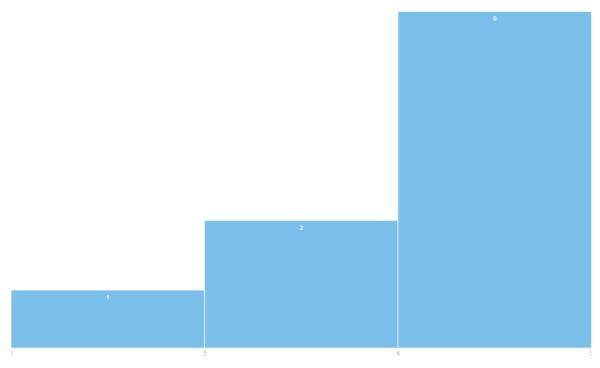
Categoria: Breweries

Anexo 15:



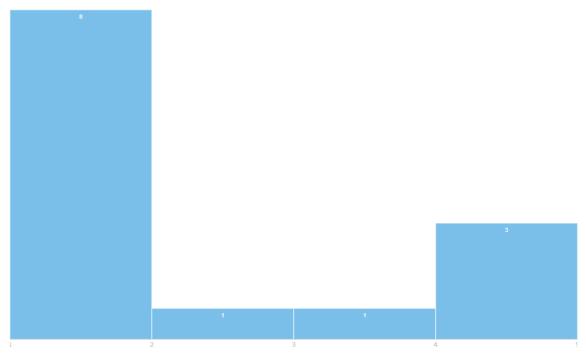
Categoria: Burgers

Anexo 16:



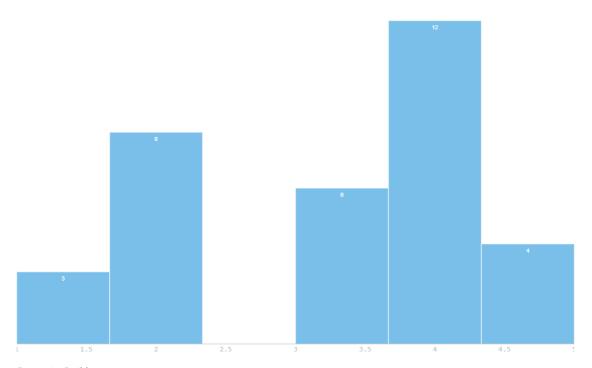
Categoria: Cafes

Anexo 17:



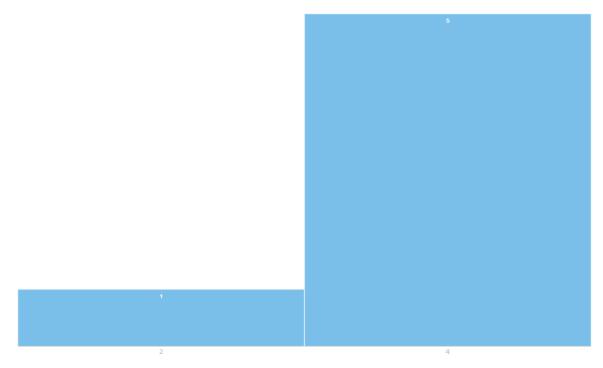
Categoria: Car Dealers

Anexo 18:



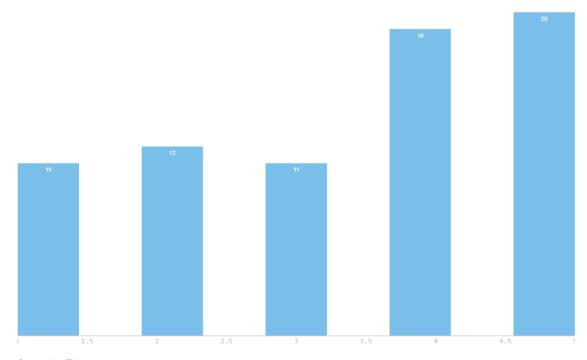
Categoria: Caribbean

Anexo 19:



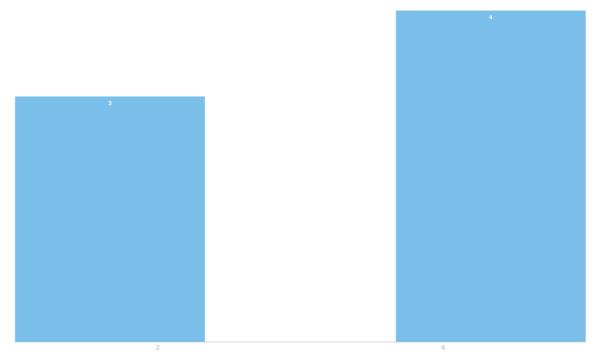
Categoria: Child Care & Day Care

Anexo 20:



Categoria: Chinese

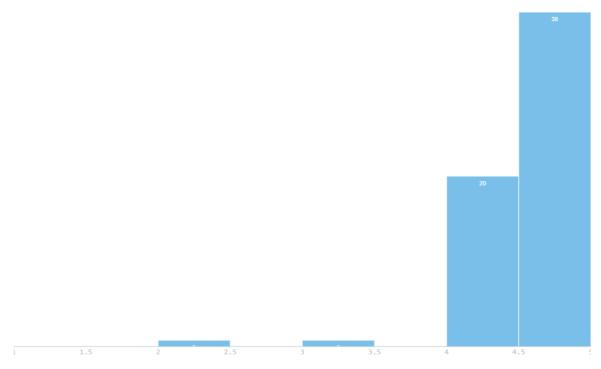
Anexo 21:



Categoria: Cocktail Bars

Anexo 22:

Anexo 23:



Categoria: Coffee & Tea

Anexo 24:

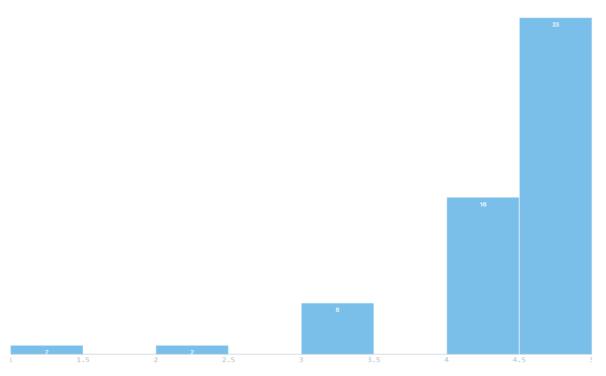


Categoria: Contractors

Anexo 25:

Categoria: Cosmetic Dentists

Anexo 26:

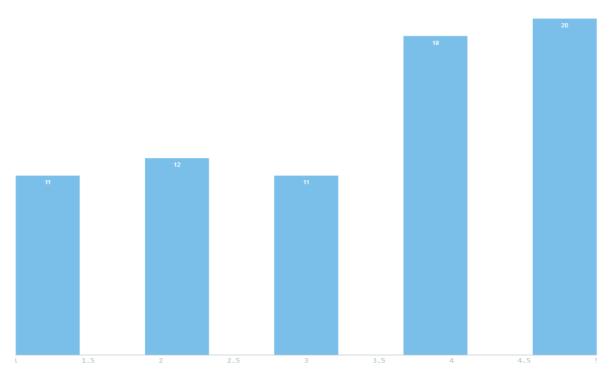


Categoria: Delis

Anexo 27:

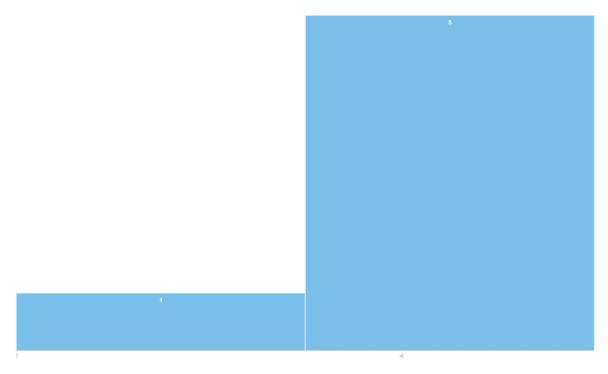
Categoria: Dentists

Anexo 28:



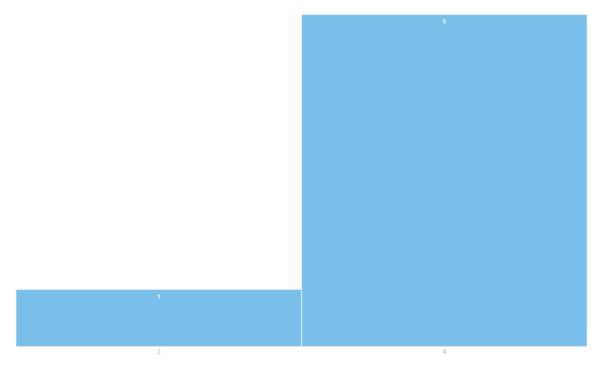
Categoria: Diners

Anexo 29:



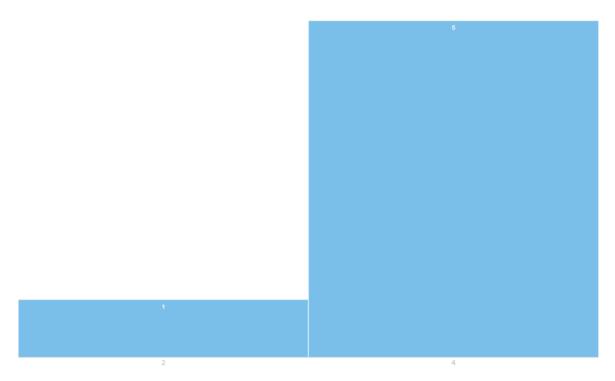
Categoria: Drugstores

Anexo 30:



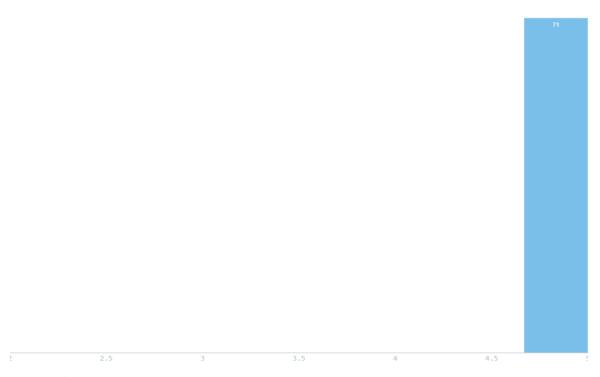
Categoria: Education

Anexo 31:



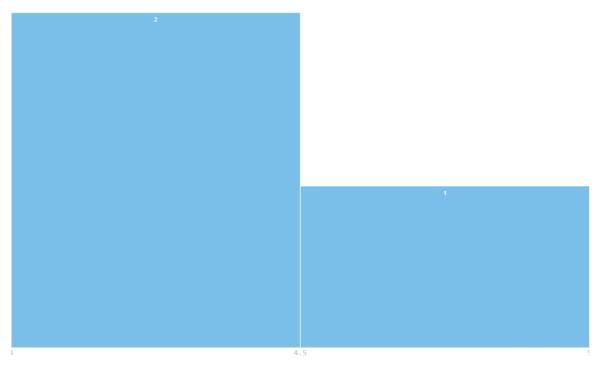
Categoria: Educational Services

Anexo 32:



Categoria: Electronics

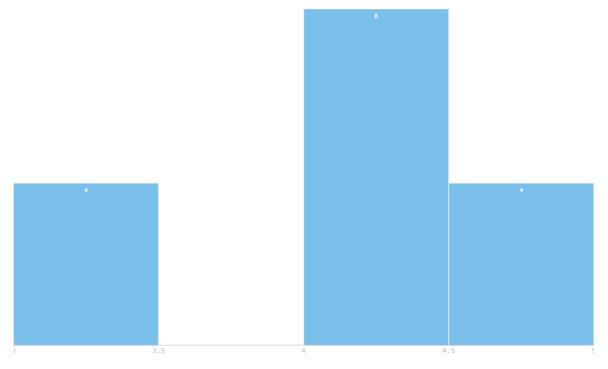
Anexo 33:



Categoria: Elementary Schools

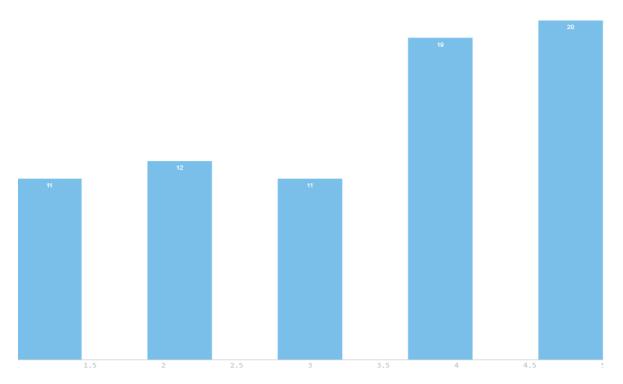
Anexo 34:

Anexo 35:



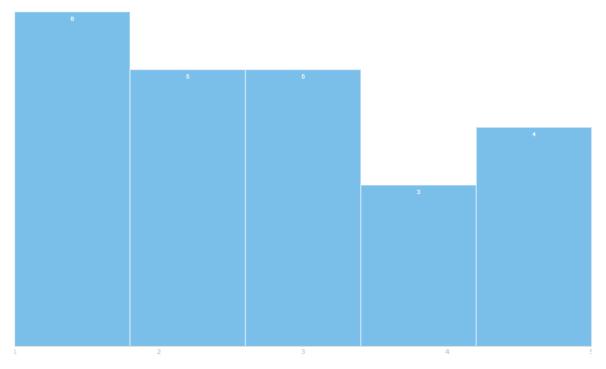
Categoria: Event Planning & Services

Anexo 36:



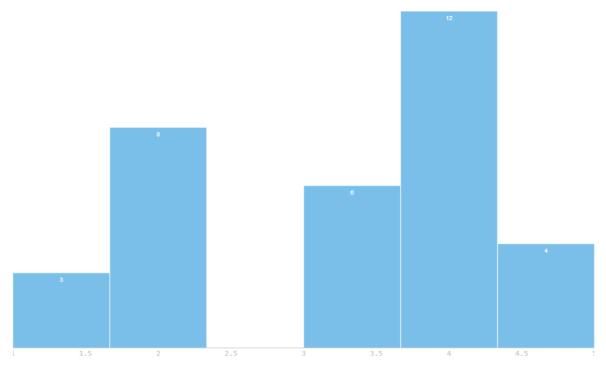
Categoria: Fast Food

Anexo 37:



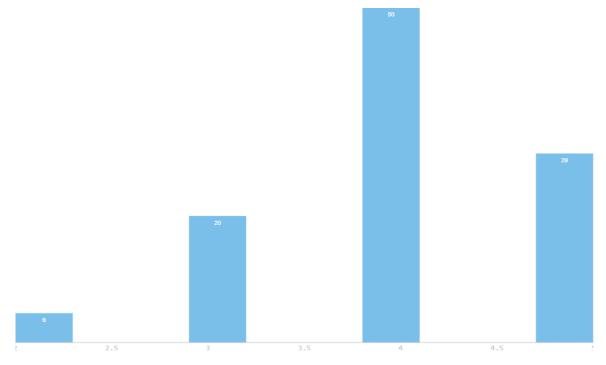
Categoria: Fish & Chips

Anexo 38:



Categoria: Food

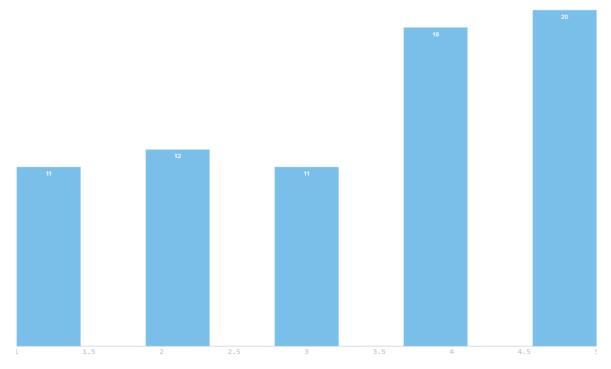
Anexo 39:



Categoria: French

Anexo 40:

Anexo 41:



Categoria: Gluten-Free

Anexo 42:



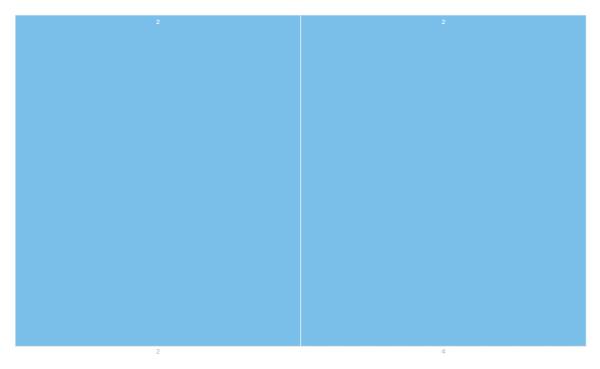
Categoria: Health & Medical

Anexo 44:



Categoria: Home Automation

Anexo 45:



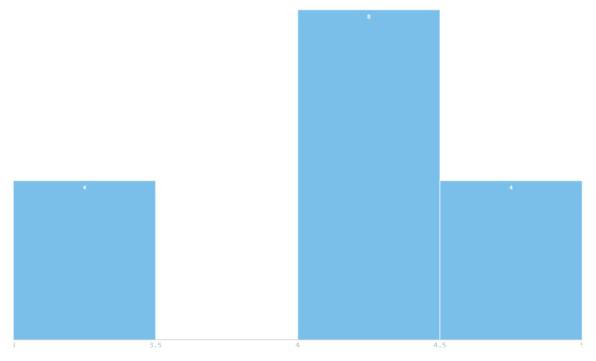
Categoria: Home Services

Anexo 46:



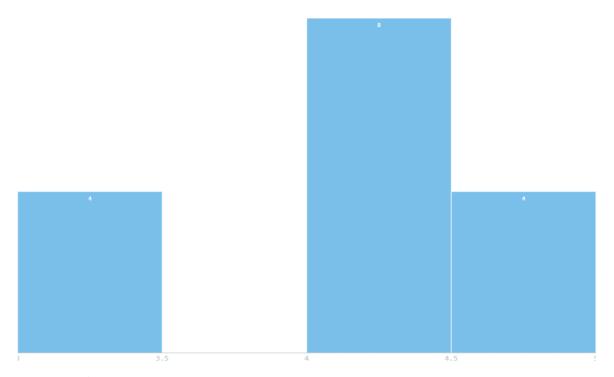
Categoria: Home Theatre Installation

Anexo 47:



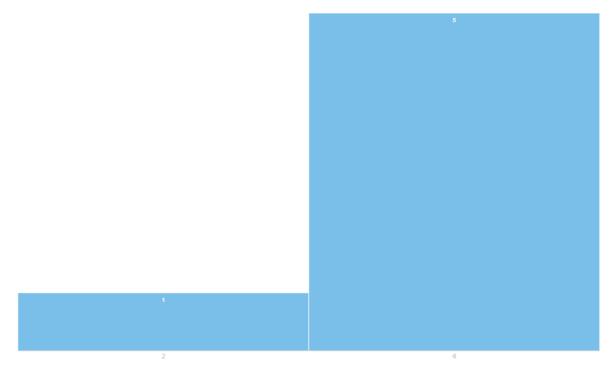
Categoria: Hotels & Travel

Anexo 48:



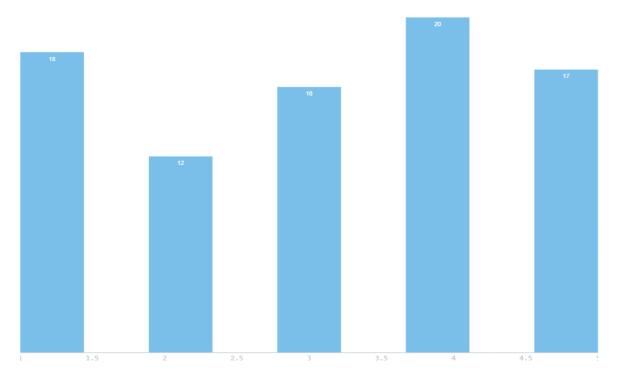
Categoria: Hotels

Anexo 49:



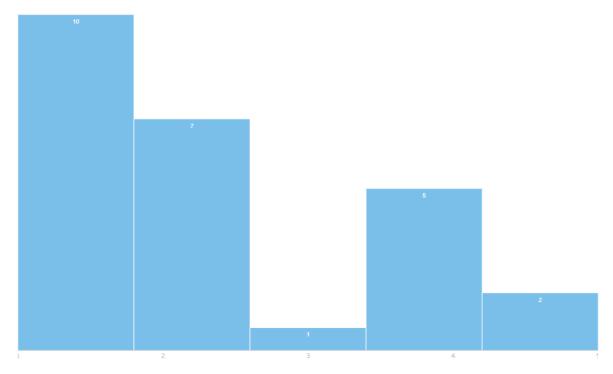
Categoria: Ice Cream & Frozen Yogurt

Anexo 50:



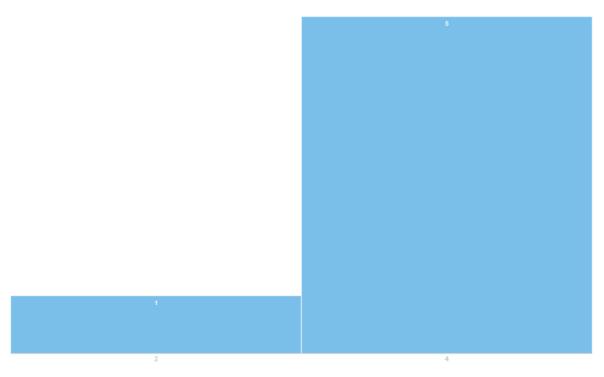
Categoria: Indian

Anexo 51:



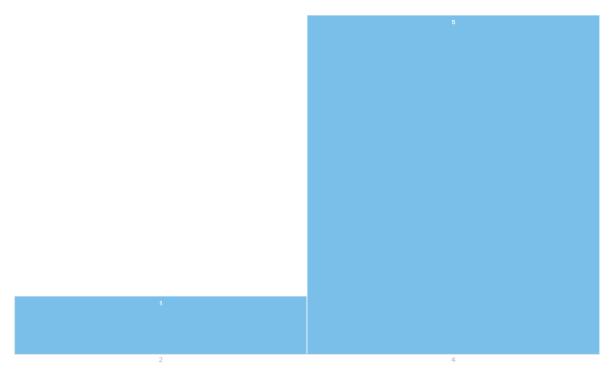
Categoria: Italian

Anexo 52:



Categoria: Language Schools

Anexo 53:



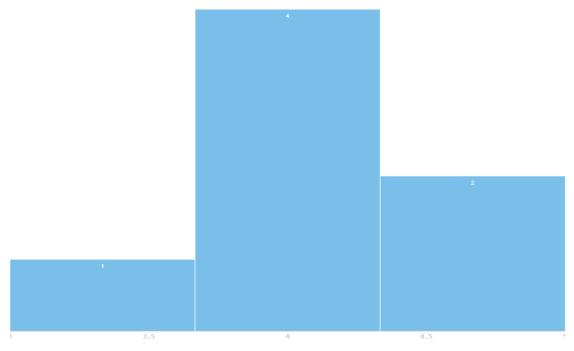
Categoria: Local Services

Anexo 54:

Anexo 55:

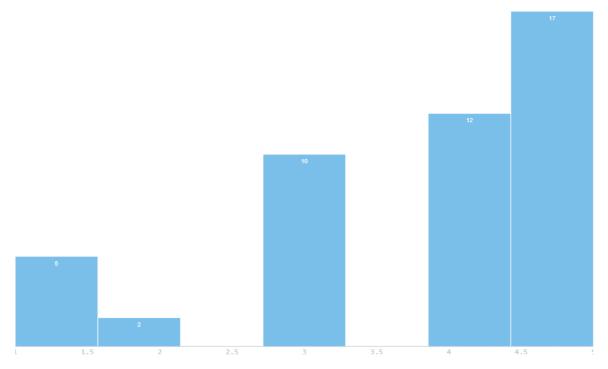
Categoria: Massage

Anexo 56:



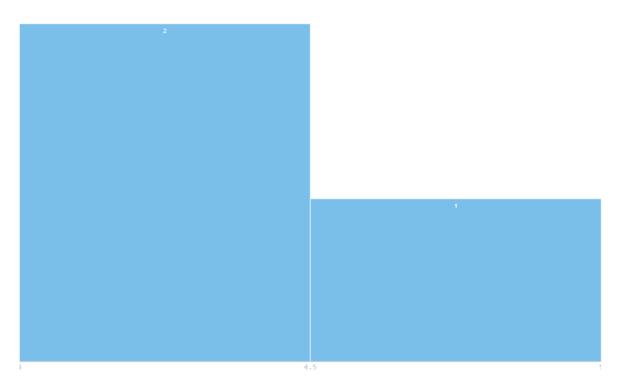
Categoria: Meat Shops

Anexo 57:



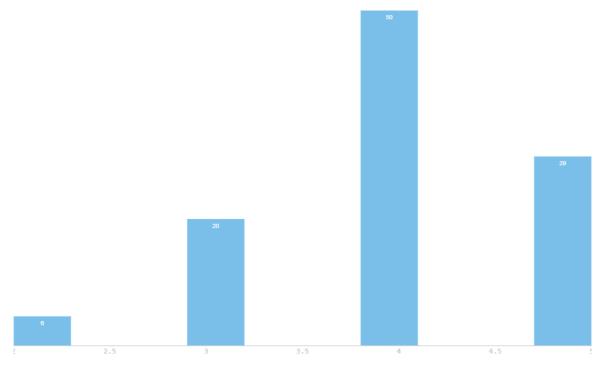
Categoria: Mexican

Anexo 58:



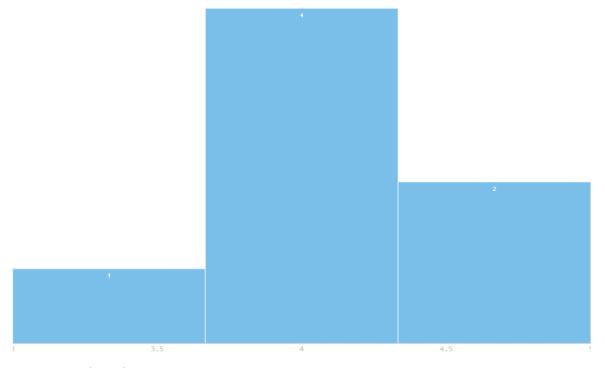
Categoria: Middle Schools & High Schools

Anexo 59:



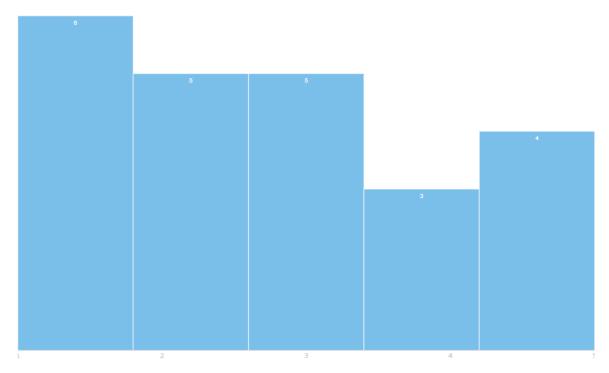
Categoria: Modern European

Anexo 60:



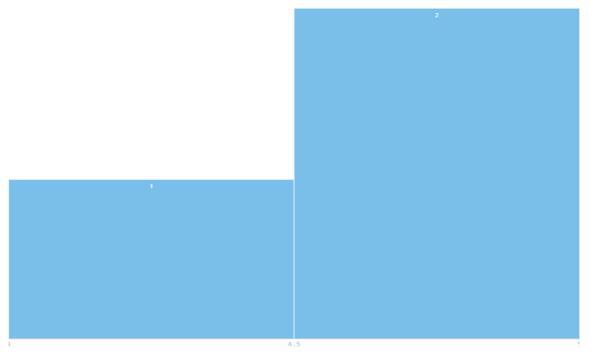
Categoria: Specialty Food

Anexo 61:



Categoria: Nightlife

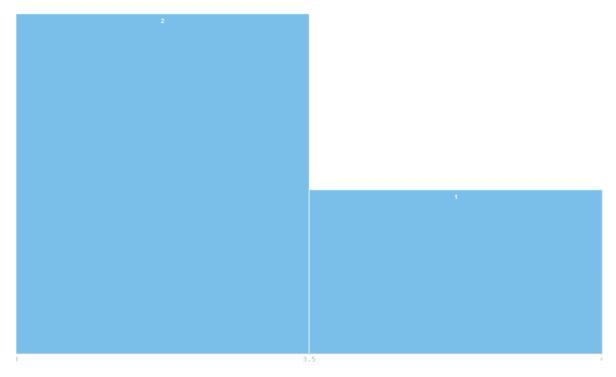
Anexo 62:



Categoria: Performing Arts

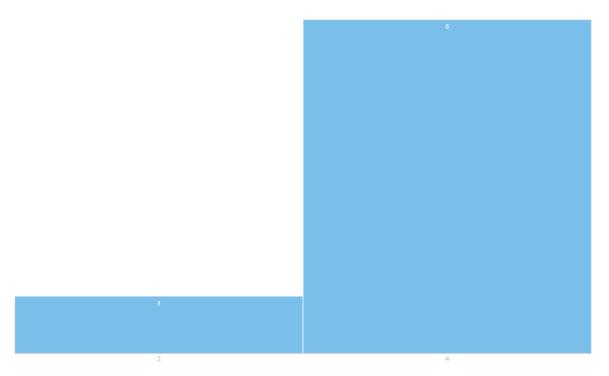
Anexo 63:		
Categoria: Periodontists		
Anexo 64:		

Anexo 65:



Categoria: Pool Halls

Anexo 66:

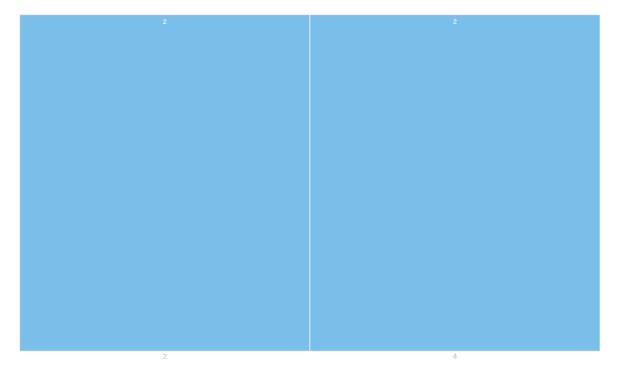


Categoria: Preschools

Anexo 67:

Categoria: Professional Services

Anexo 68:

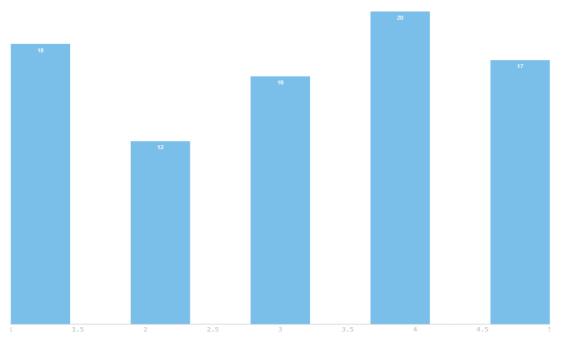


Categoria: Real Estate

Anexo 69:

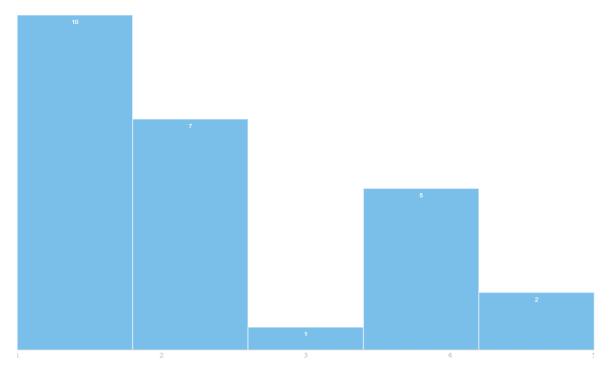
Categoria: Reiki

Anexo 70:



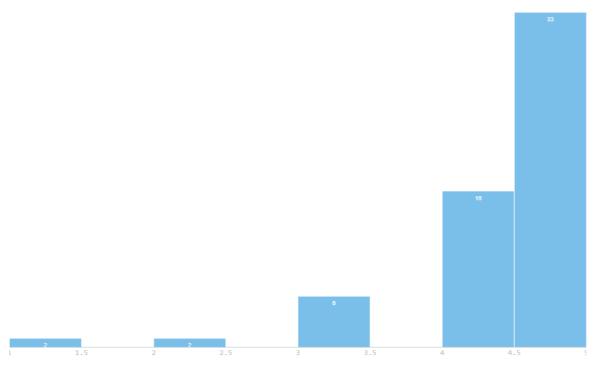
Categoria: Restaurants

Anexo 71:



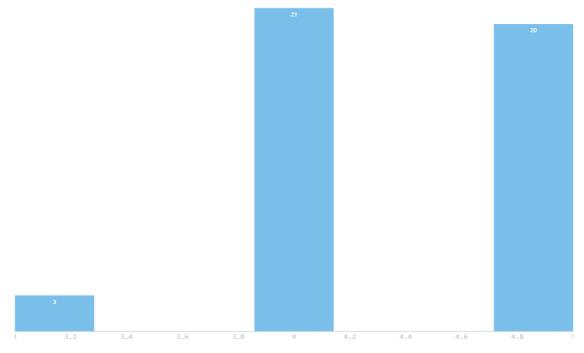
Categoria: Salad

Anexo 72:



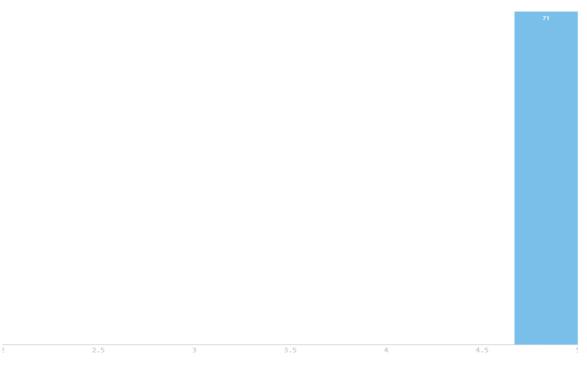
Categoria: Sandwiches

Anexo 73:



Categoria: Seafood

Anexo 74:

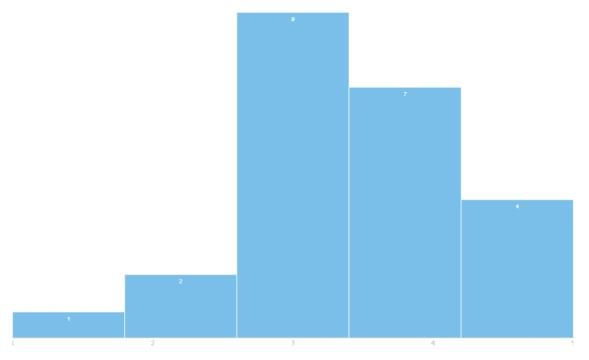


Categoria: Security Systems

Anexo 75:

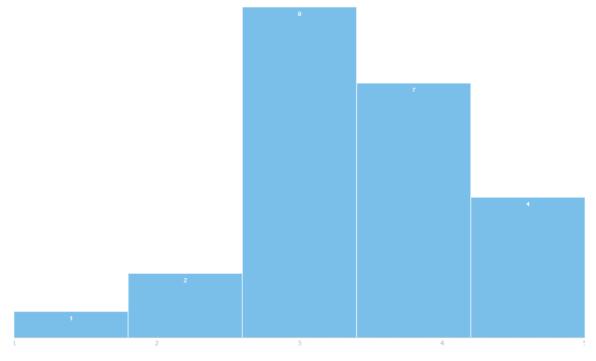
Categoria: Session Photography

Anexo 76:



Categoria: Shopping Centers

Anexo 77:



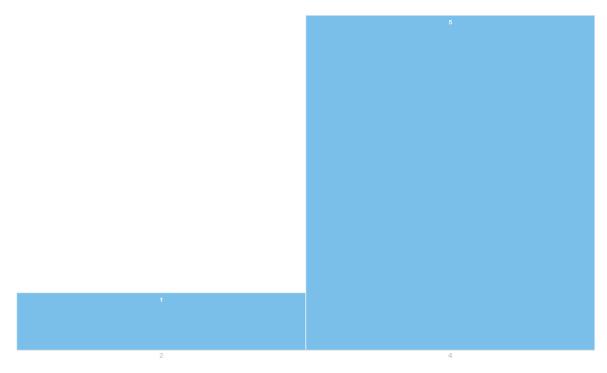
Categoria: Shopping

Anexo 78:



Categoria: Soul Food

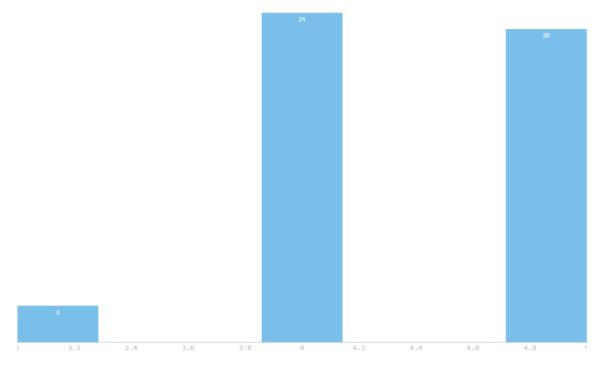
Anexo 79:



Categoria: Specialty Schools

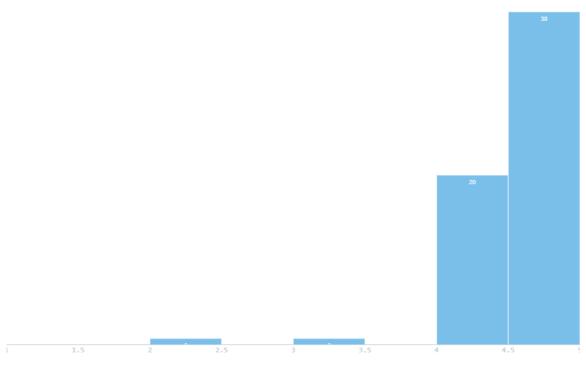
Anexo 80:

Anexo 81:



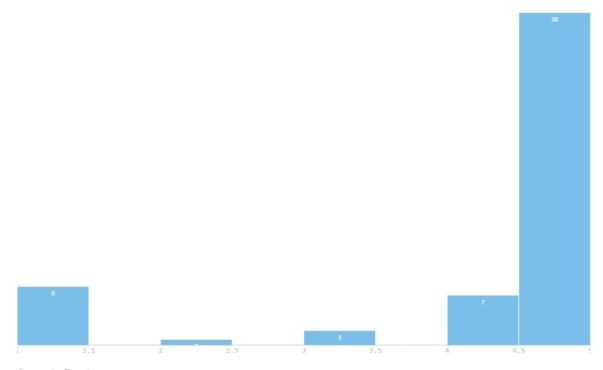
Categoria: Steakhouses

Anexo 82:



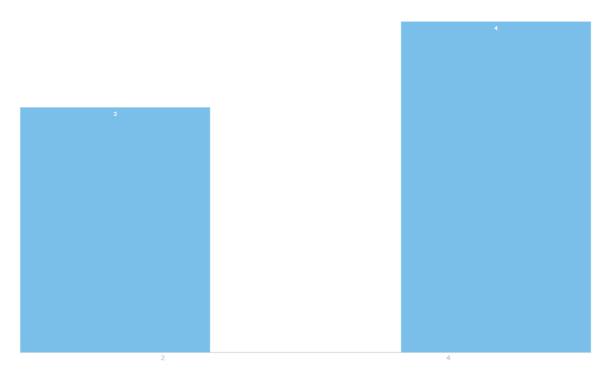
Categoria: Sushi Bars

Anexo 83:



Categoria: Tanning

Anexo 84:

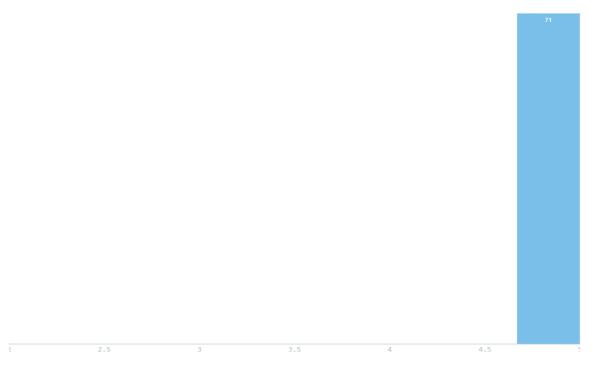


Categoria: Tapas Bars

Anexo 85:

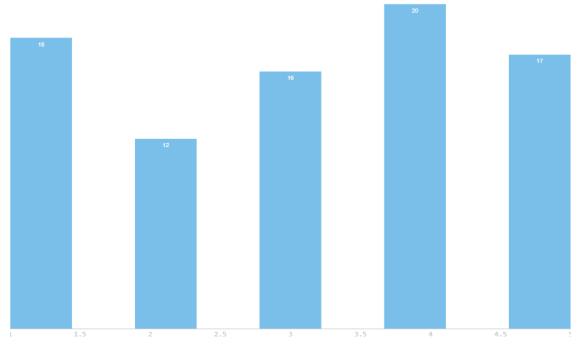
Categoria: Tours

Anexo 86:



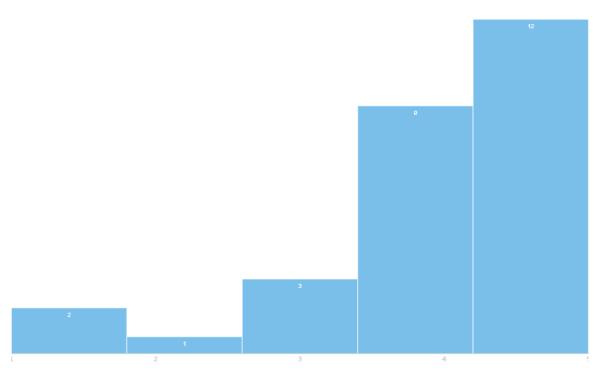
Categoria: TV Mounting

Anexo 87:



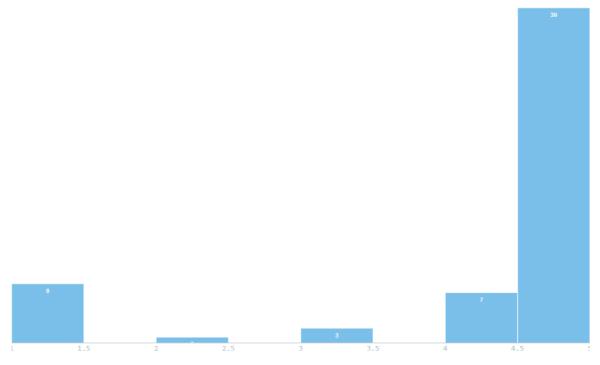
Categoria: Vegetarian

Anexo 88:



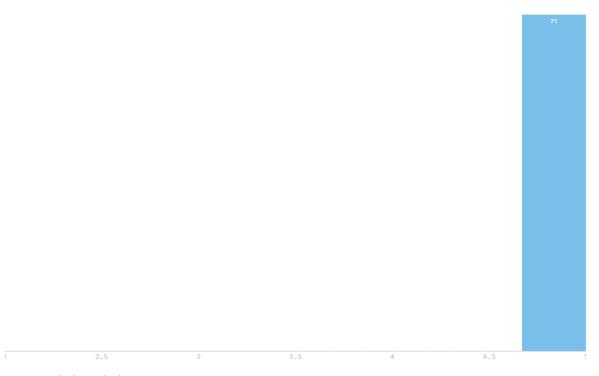
Categoria: Japanese

Anexo 89:



Categoria: Nail Salons

Anexo 90:



Categoria: Shades & Blinds