IRONHACK - DATA

DIAMONDS DATA

Antonio Huerta

PRESENTACIÓN

DIAMONDS DATA

- Proceso & Stack
- Ejemplos
- Output
- Conclusiones

Funcionamiento del proyecto

1

Data Extraction

We extract data with DBeaber. We jpin table in a single sataset

Diamonds_db_SQL.csv

2

Data Wrangling

We add continent column using pandas

Diamonds_db_pandas.csv

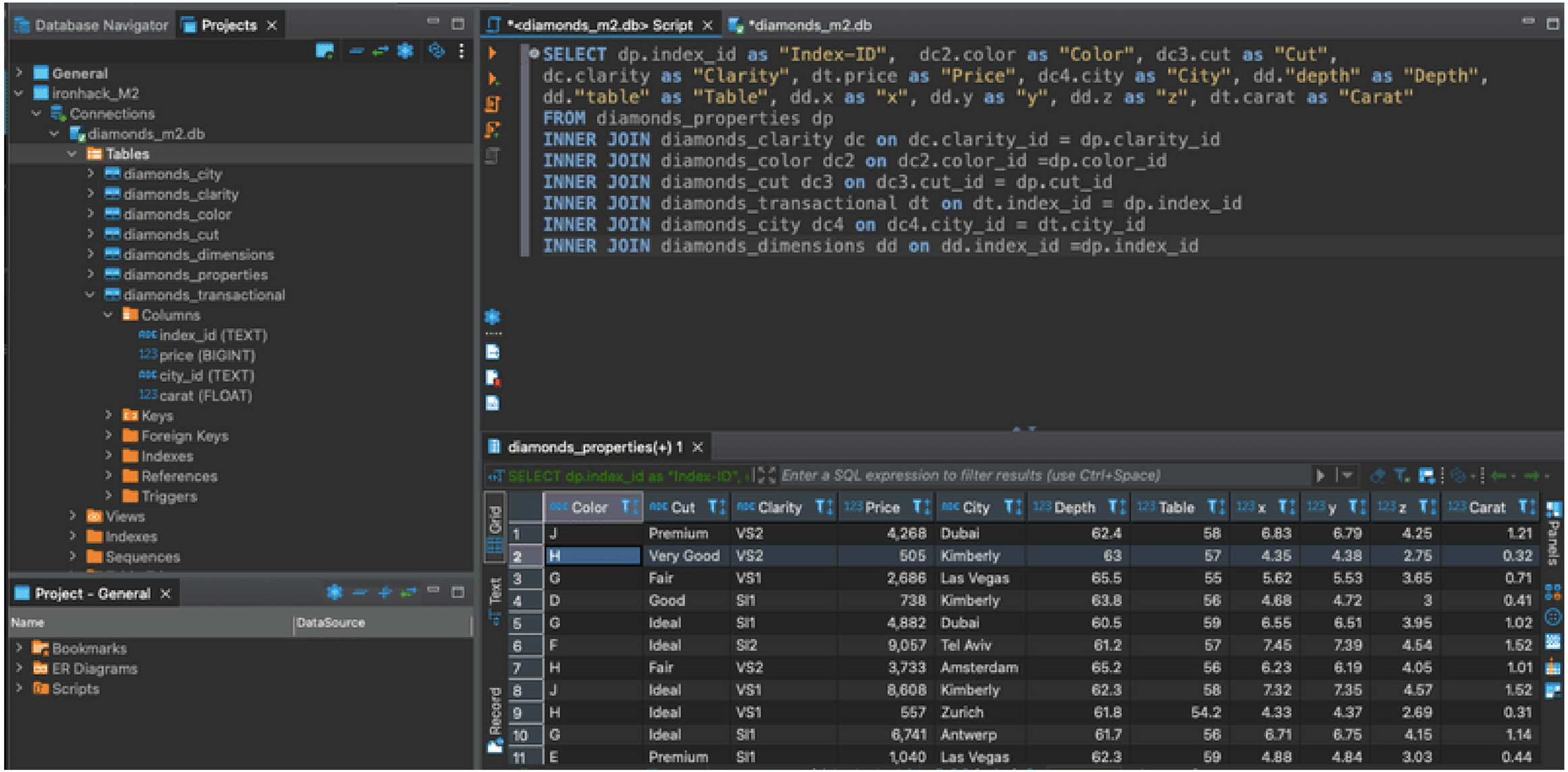
3

Visualization

ShowMybike will save a file in your data folder called "location_one.csv" or "location_all.csv"

Public Tableau

Diamonds_db_SQL



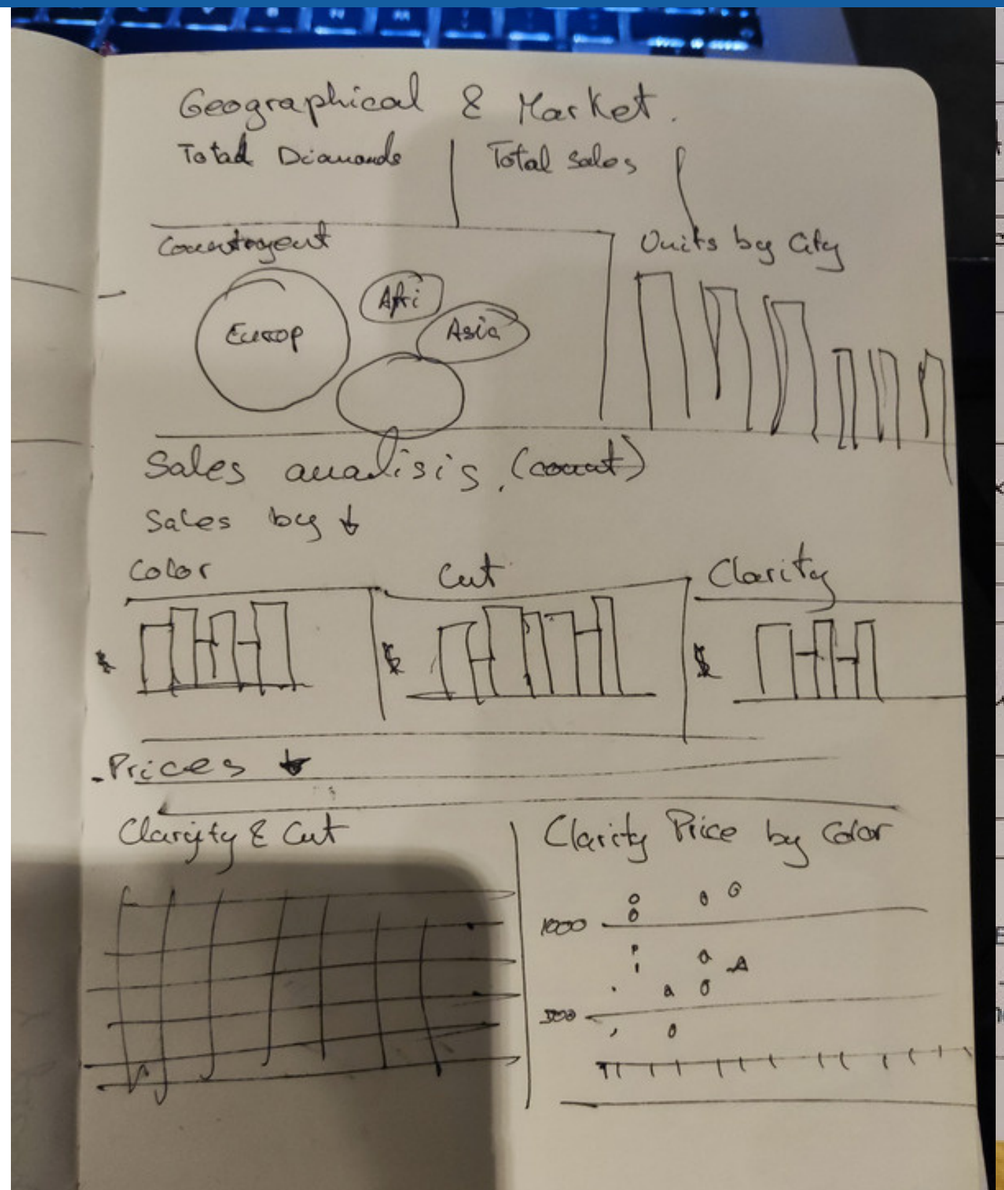
	Color	Cut	Clarity	Price	City	Depth	Table	x	y	z	Carat
1	J	Premium	VS2	4,268	Dubai	62.4	58	6.83	6.79	4.25	1.21
2	H	Very Good	VS2	505	Kimberly	63	57	4.35	4.38	2.75	0.32
3	G	Fair	VS1	2,686	Las Vegas	65.5	55	5.62	5.53	3.65	0.71
4	D	Good	SI1	738	Kimberly	63.8	56	4.68	4.72	3	0.41
5	G	Ideal	SI1	4,882	Dubai	60.5	59	6.55	6.51	3.95	1.02
6	F	Ideal	SI2	9,057	Tel Aviv	61.2	57	7.45	7.39	4.54	1.52
7	H	Fair	VS2	3,733	Amsterdam	65.2	56	6.23	6.19	4.05	1.01
8	J	Ideal	VS1	8,608	Kimberly	62.3	58	7.32	7.35	4.57	1.52
9	H	Ideal	VS1	557	Zurich	61.8	54.2	4.33	4.37	2.69	0.31
10	G	Ideal	SI1	6,741	Antwerp	61.7	56	6.71	6.75	4.15	1.14
11	E	Premium	SI1	1,040	Las Vegas	62.3	59	4.88	4.84	3.03	0.44

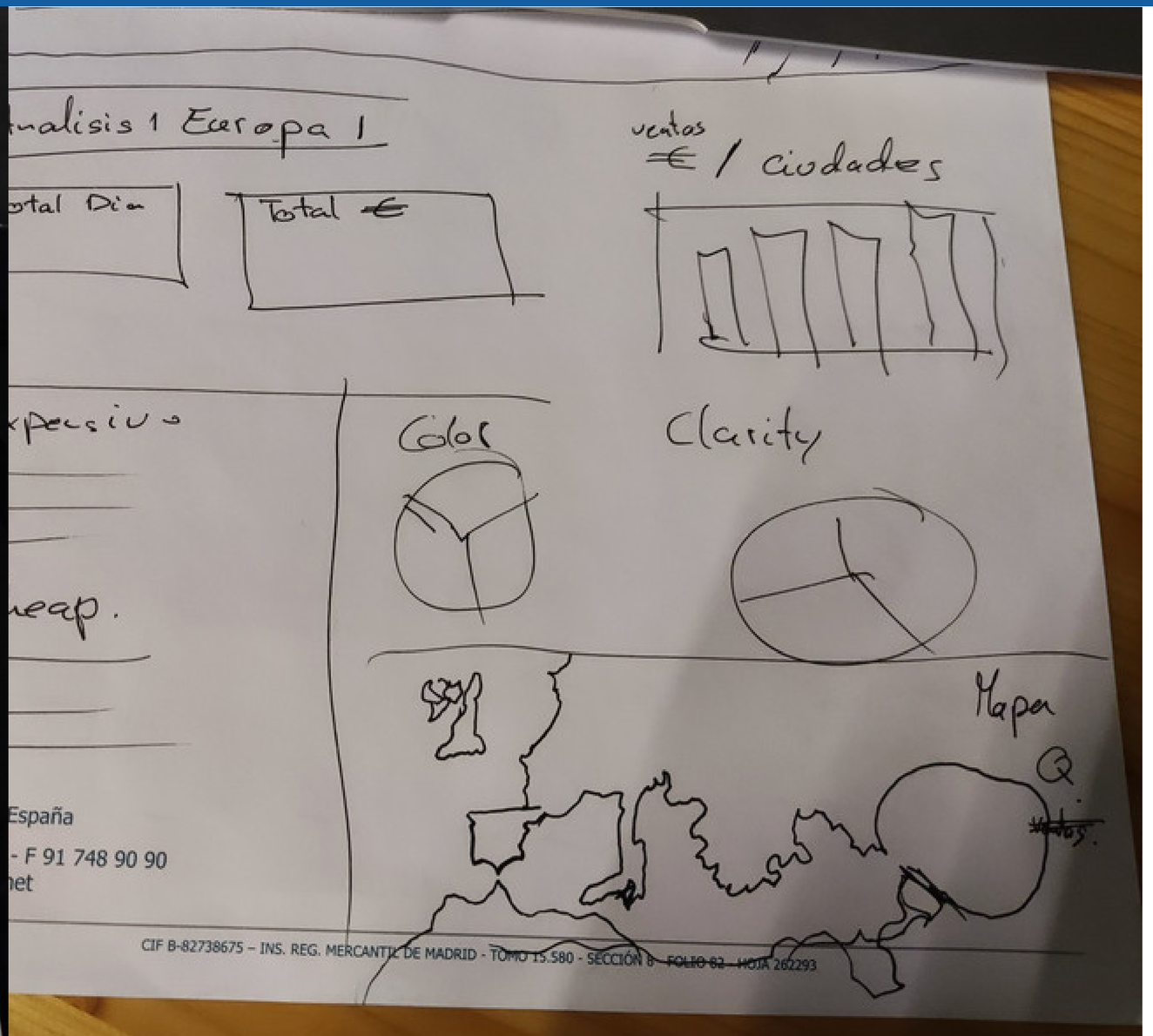
continents.py

Pandas_modules > continents.py > ...

```python
# Libraries
import pandas as pd
import numpy as np

# Import CSV from SQL
df_diamonds = pd.read_csv ('/Users/antoniohuerta/ironhack/ih_datamadpt1121_project_m2/db/Diamonds_db_SQL.csv')

# Obtenemos el listado de valores únicos de la columna city
cities = df_diamonds['City'].unique()

# Creamos un diccionario al que asignamos los continentes:

country_cont = {'Dubai':'Asia',
                'Kimberly':'Africa',
                'Las Vegas':'America',
                'Tel Aviv':'Asia',
                'Amsterdam':'Europe',
                'Zurich':'Europe',
                'Antwerp':'Europe',
                'Madrid':'Europe',
                'Paris':'Europe',
                'Surat':'Asia',
                'Luxembourg':'Europe',
                'London':'Europe',
                'New York City':'America'
                }

# Insertamos una columna con el continente
df_diamonds['Continent']=df_diamonds['City'].map(country_cont)
# Exportamos el csv
df_diamonds.to_csv(r'/Users/antoniohuerta/ironhack/ih_datamadpt1121_project_m2/db/Diamonds_db_pandas.csv', index = False)
```

Sketching


Geographical & Market.
Total Diamonds
Total sales
Units by City
Sales analisis (count)
Sales by
Color
Cut
Clarity
Prices
Clarity & Cut
Clarity Price by Color


Total €
ventas
€ / ciudades
Color
Clarity
Mapa
España
- F 91 748 90 90
CIF B-82738675 – INS. REG. MERCANTIL DE MADRID - TOMO 15.580 - SECCIÓN 8 - FOLIO 82 - HOJA 262293

Tableau

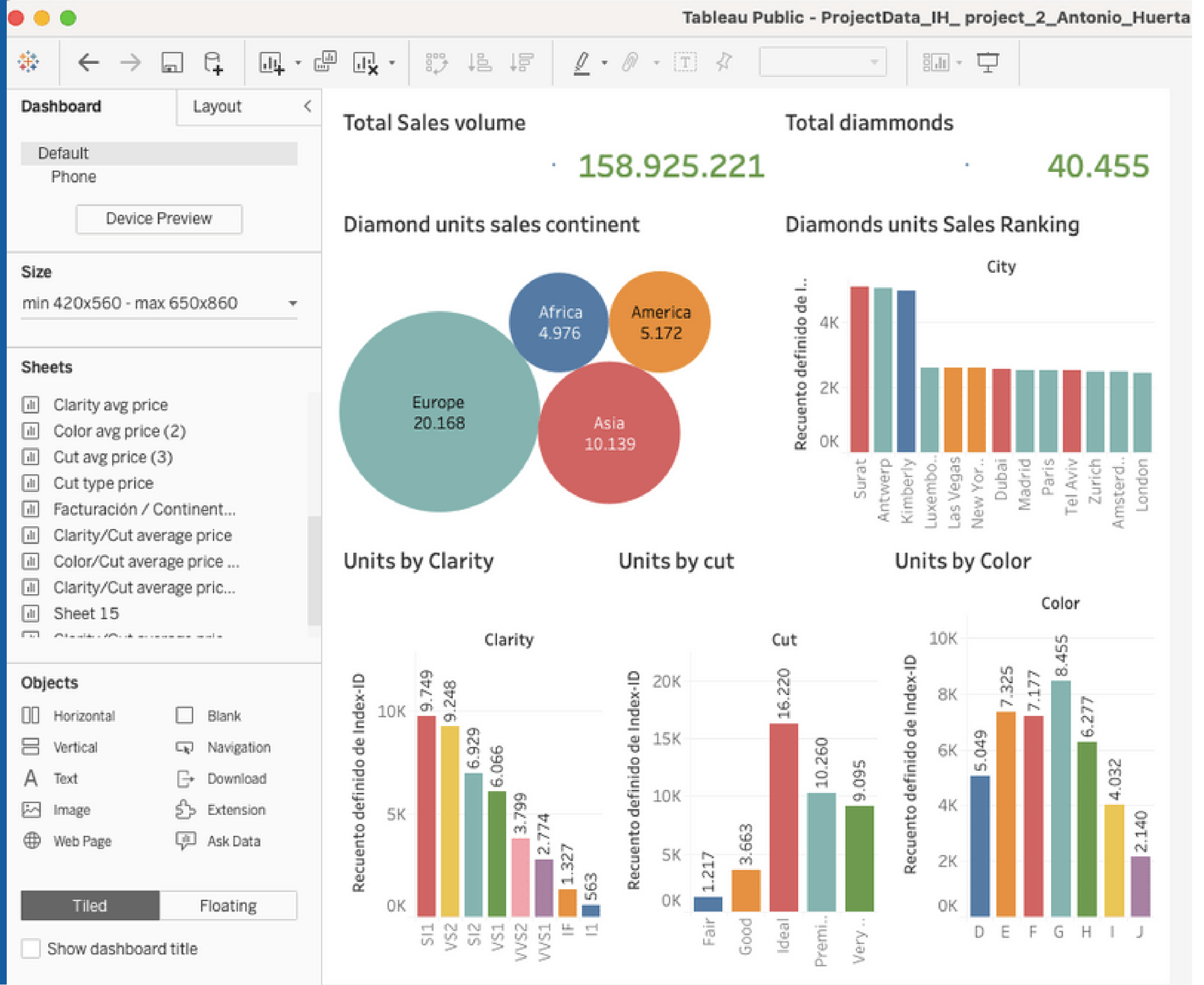


Tableau Public - ProjectData_IH_ project_2_Antonio_Huerta
Dashboard
Layout
Default
Phone
Device Preview
Size
min 420x560 - max 650x860
Sheets
Clarity avg price
Color avg price (2)
Cut avg price (3)
Cut type price
Facturación / Continent...
Clarity/Cut average price
Color/Cut average price ...
Clarity/Cut average pric...
Sheet 15
Objects
Horizontal
Blank
Vertical
Navigation
Text
Download
Image
Extension
Web Page
Ask Data
Tiled
Floating
Show dashboard title
Total Sales volume
158.925.221
Total diammonds
40.455
Diamond units sales continent
Africa
4.976
America
5.172
Europe
20.168
Asia
10.139
Diamonds units Sales Ranking
City
Recuento definido de I..
4K
2K
0K
Surat
Antwerp
Kimberly
Luxembo..
Las Vegas
New Yor..
Dubai
Madrid
Paris
Tel Aviv
Zurich
Amsterd..
London
Units by Clarity
Units by cut
Units by Color
Clarity
Recuento definido de Index-ID
10K
5K
0K
9.749
9.248
6.929
6.066
3.799
2.774
1.327
563
SI1
VS2
SI2
VS1
VVS2
VVS1
IF
I1
Cut
20K
15K
10K
5K
0K
1.217
3.663
16.220
10.260
9.095
Fair
Good
Ideal
Premi..
Very ..
Color
10K
8K
6K
4K
2K
0K
5.049
7.325
7.177
8.455
6.277
4.032
2.140
D
E
F
G
H
I
J

Conclusiones Dashboard

GEOGRAPICAL DISTRIBUTION

- En Europa se realizan el doble de transacciones que en Asia, y cuatro veces más que África y Asia.
- Surat, Amberes y Kimberly son las tres ciudades con mayor volumen de ventas, más del doble que las 10 restantes.

CLARITY, CUT, COLOR

- SI1 es el Clarity que más ventas tiene con un total de 9747 uds, tan sólo 500 unidades más que VS2.
- El color G es el más demandado con 8455 uds, seguido de cerca por el D y el F con 7325 y 7177 respectivamente.
- Cut "Ideal" es el más demandado por los usuarios.

AVERAGE PRICES

Dentro del corte "Ideal", Los Color D tienen un precio medio más bajo (2588) seguidos de E (2625) y F (3284).

Por otro lado el Cut "Premium" tiene los precios medios más altos en sus variantes de color J (6377) e I (6023)

CRITERIOS DE PROPIEDAD

El color D tiene los precios medios más altos excepto en las variedades de corte "Ideal" y "Fair".

La combinación de Clarity S12 y Color I es el valor más estable en precios medios, independientemente de su Cut

¡GRACIAS!

IRONHACK - ANTONIO HUERTA