```
print·("Olá·Mundo!·Projeto·01·Salários·de·trabalho·em·ciência·de·dados")
```

```
Olá Mundo! Projeto 01 Salários de trabalho em ciência de dados
```

```
#Instalando bibliotéca SWEETVIZ de visualização.
pip install sweetviz
```

```
Looking in indexes: https://pypi.org/simple, https://us-python.pkg.dev/colab-wheels/
Collecting sweetviz
  Downloading sweetviz-2.1.4-py3-none-any.whl (15.1 MB)
     |████████████████████████████████| 15.1 MB 5.1 MB/s
Requirement already satisfied: matplotlib>=3.1.3 in /usr/local/lib/python3.7/dist-pa
Requirement already satisfied: pandas!=1.0.0,!=1.0.1,!=1.0.2,>=0.25.3 in /usr/local/
Requirement already satisfied: jinja2>=2.11.1 in /usr/local/lib/python3.7/dist-packa
Requirement already satisfied: tqdm>=4.43.0 in /usr/local/lib/python3.7/dist-package
Requirement already satisfied: numpy>=1.16.0 in /usr/local/lib/python3.7/dist-packag
Requirement already satisfied: scipy>=1.3.2 in /usr/local/lib/python3.7/dist-package
Requirement already satisfied: importlib-resources>=1.2.0 in /usr/local/lib/python3.
Requirement already satisfied: zipp>=3.1.0 in /usr/local/lib/python3.7/dist-packages
Requirement already satisfied: MarkupSafe>=0.23 in /usr/local/lib/python3.7/dist-pac
Requirement already satisfied: kiwisolver>=1.0.1 in /usr/local/lib/python3.7/dist-pa
Requirement already satisfied: pyparsing!=2.0.4,!=2.1.2,!=2.1.6,>=2.0.1 in /usr/loca
Requirement already satisfied: cycler>=0.10 in /usr/local/lib/python3.7/dist-package
Requirement already satisfied: python-dateutil>=2.1 in /usr/local/lib/python3.7/dist
Requirement already satisfied: typing-extensions in /usr/local/lib/python3.7/dist-pa
Requirement already satisfied: pytz>=2017.3 in /usr/local/lib/python3.7/dist-package
Requirement already satisfied: six>=1.5 in /usr/local/lib/python3.7/dist-packages (f
Installing collected packages: sweetviz
Successfully installed sweetviz-2.1.4
```

```
#Importando bibliotéca SWEETVIZ de visualização.
import sweetviz as sz
```

```
#Importando Biblioteca PANDAS para manitulação dos dados
import pandas as pd
```

```
#Importando arquivo CSV com Pandas
var1 = pd.read_csv('ds_salaries.csv')
```

```
#Ver dados importado na VAR1
var1.head(n=5)
```

1 to 5 of 5 entries Filter

index	Unnamed: 0	work_year	experience_level	employment_type	job_title	salary	salary_c
0	0	2020	MI	FT	Data Scientist	70000	EUR
1	1	2020	SE	FT	Machine Learning Scientist	260000	USD
2	2	2020	SE	FT	Big Data Engineer	85000	GBP

```
#Detalhes da tabela importada
var1.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 607 entries, 0 to 606
Data columns (total 12 columns):
 #   Column              Non-Null Count  Dtype 
---  ------              --------------  ----- 
 0   Unnamed: 0          607 non-null    int64 
 1   work_year           607 non-null    int64 
 2   experience_level    607 non-null    object
 3   employment_type     607 non-null    object
 4   job_title           607 non-null    object
 5   salary              607 non-null    int64 
 6   salary_currency     607 non-null    object
 7   salary_in_usd       607 non-null    int64 
 8   employee_residence  607 non-null    object
 9   remote_ratio        607 non-null    int64 
 10  company_location    607 non-null    object
 11  company_size        607 non-null    object
dtypes: int64(5), object(7)
memory usage: 57.0+ KB
```

```
#Informações Estatísticas das linhss númerias: Count=contagem, Mean=média,
#std=desvio padrão, mim=mínimo.
var1.describe()
```

	Unnamed: 0	work_year	salary	salary_in_usd	remote_ratio
count	607.000000	607.000000	6.070000e+02	607.000000	607.00000
mean	303.000000	2021.405272	3.240001e+05	112297.869852	70.92257
std	175.370085	0.692133	1.544357e+06	70957.259411	40.70913
min	0.000000	2020.000000	4.000000e+03	2859.000000	0.00000
25%	151.500000	2021.000000	7.000000e+04	62726.000000	50.00000
50%	303.000000	2022.000000	1.150000e+05	101570.000000	100.00000
75%	454.500000	2022.000000	1.650000e+05	150000.000000	100.00000
max	606.000000	2022.000000	3.040000e+07	600000.000000	100.00000

```
#Correlações entre as variaveis da tabela.
var1.corr()
```

1 to 5 of 5 entries Filter

index	Unnamed: 0	work_year	salary	salary
Unnamed: 0	1.0	0.8865498824882385	-0.09624957831171785	0.1670246
work_year	0.8865498824882385	1.0	-0.0875773838945656	0.1704933
salary	-0.09624957831171785	-0.0875773838945656	1.0	-0.083905
salary_in_usd	0.16702467218076655	0.17049332455338606	-0.0839056921059893	
remote_ratio	0.09499961910471186	0.07631437405384806	-0.014608470101377384	0.1321223

Show 25 per page

Like what you see? Visit the data table notebook to learn more about interactive tables.

```
#visualização dos dados
relatorio = sz.analyze(var1)
```

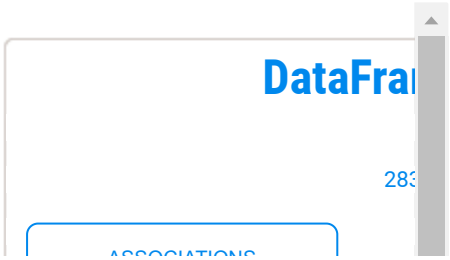
Done! Use 'show' commands to display/save. [100%] 00:01 -> (00:00 left)

```
relatorio.show_notebook()
```

2.1.4

Get updates, docs & report issues here

Created & maintained by Francois Bertrand



```
#Outra biblipteca de visualização caso precisar. "pip install seaborn"
```

DataFrame

```
#Fonte dos dados: https://www.kaggle.com/datasets/ruchi798/data-science-job-salaries/discu
```

VALUES:	607 (100%)	MAX	606	RANGE
MISSING:	---	95%	576	IQR
		Q3	454	STD
DISTINCT:	607 (100%)	MEDIAN	303	VAR
		AVG	303	
ZEROES:	1 (<1%)	Q1	152	KURT.
		5%	30	SKEW
		MIN	0	SUM

2 work_year

VALUES:	607 (100%)
MISSING:	---
DISTINCT:	3 (<1%)

3 experience_level

VALUES:	607 (100%)
MISSING:	---
DISTINCT:	4 (<1%)

✓ 0s conclusão: 12:54