

Other people's Data

(and how to not hate it)

Raise your hand
if you've ever
hated
a dataset.

Raise your hand
if it was data
you
collected.



Mali Akmanalp

@makmanalp

THE ATLAS OF ECONOMIC COMPLEXITY

MAPPING PATHS TO PROSPERITY

Ricardo Hausmann, César A. Hidalgo, et al.



“lol noone uses python for stats, numpy indexing is awful and its a terrible language for data analysis. get out.” (sic)

–Economist e86d on econjobrumors.com

Live Demo

`read_csv`
`read_excel`
`read_hdf`
`read_sql`
`read_json`
`read_msgpack`
`read_html`
`read_gbq`
`read_stata`
`read_clipboard`
`read_pickle`

I went to the cafÃ© this morning

unicodedata / file

Get the right column type!

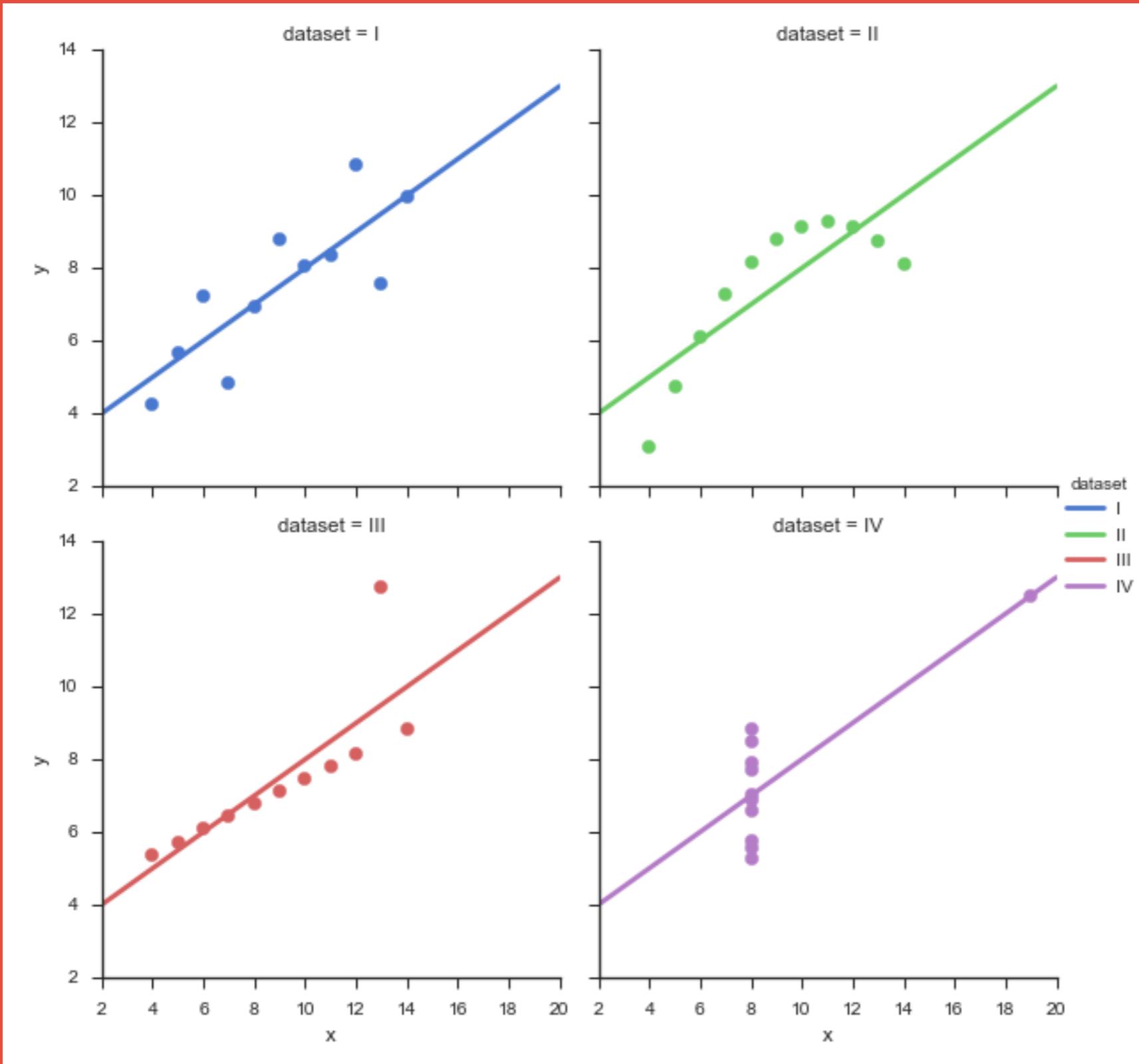
int vs float

object vs string

Get an overview!

.info()
.describe()

Does the data make sense?



Seaborn
ggplot
statsmodels

Which things in this column are not unique?
`duplicated()`

How many of each thing do I have?
`value_counts()`

Filling empty data

```
Series.interpolate(method='linear', axis=0, limit=None,  
inplace=False, downcast=None, **kwargs)
```

method : {‘linear’, ‘time’, ‘index’, ‘values’, ‘nearest’, ‘zero’,
‘slinear’, ‘quadratic’, ‘cubic’, ‘barycentric’, ‘krogh’, ‘polynomial’,
‘spline’ ‘piecewise_polynomial’, ‘pchip’}

Filling empty data

DataFrame.fillna(value=None, method=None, axis=0,
inplace=False, limit=None, downcast=None)

{‘backfill’, ‘bfill’, ‘pad’, ‘ffill’, None}

Merging Tables

The complement of the table on the right,
which is the result of a LEFT JOIN.

ENGLISH_TEXT	ENGLISH_ID
One	1
Two	2
Three	3
Four	4
Five	5
Six	6

FRENCH_ID	FRENCH_TEXT
1	Un
3	Trois
4	Quatre
5	Cinq
6	Six
7	Sept
8	Huit

The complement of the table on the left,
which is the result of a RIGHT JOIN.

Merging Tables

```
DataFrame.join(other, on=None, how='left', lsuffix='', rsuffix='',  
sort=False)
```

```
DataFrame.merge(right, how='inner', on=None, left_on=None,  
right_on=None, left_index=False, right_index=False,  
sort=False, suffixes=('_x', '_y'), copy=True)
```

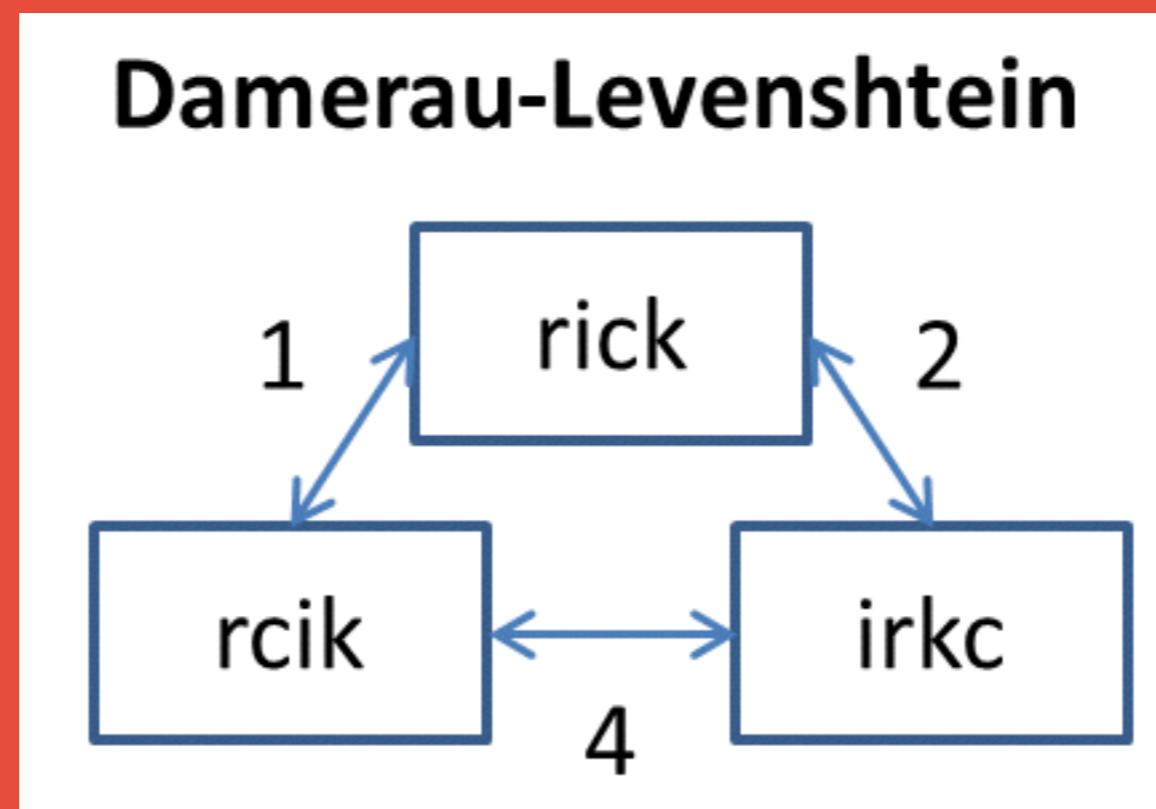
Check column sizes after merge with .count()

Merging Tables

Check column sizes after merge with .count()

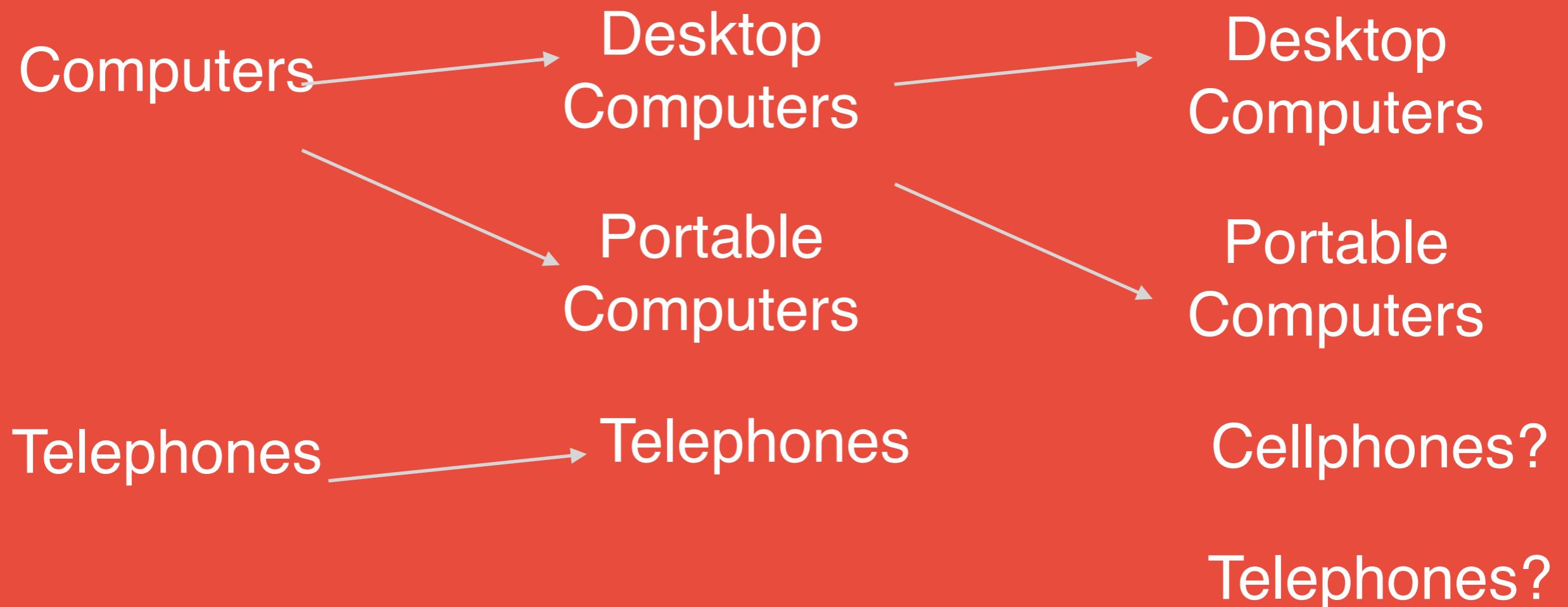
Compare with original data!

Fuzzy matching On text fields



jellyfish (Sunlight foundation)
elasticsearch

Network Visualization for concordance matches



networkx

Longer Notebooks

write small convenience functions to reduce long-line-syndrome

name things correctly, “df” and “df2” are not your friend

PDFs Suck

Tabula is experimental software Home About

Page 1

Page 2

Monthly Employment Situation for Veterans
Released Monthly on the day of the U.S. Employment Situation.
Economic Information & Analysis, Illinois Department of Employment Security

Release date: December 6, 2013

National Data [Only data from the Monthly Report, Table A-5 will be updated monthly]
The U.S. Bureau of Labor Statistics reports monthly the employment status of the civilian population 16 years and over by veteran status, period of service and sex, not seasonally adjusted. (Table A-5) In November 2013, the unemployment rate for all Veterans over 18 was 6.7%, up from 6.6% in November 2012. The unemployment rate for Gulf War-era II Veterans over 18 was 9.9% in November 2013, down from 10.0% in November 2012. These numbers are not seasonally adjusted; they come from a small sample and are highly volatile from one month to the next. The December employment figures will be available on January 10, 2014. Here is a link to the November U.S. Employment Situation http://www.bls.gov/schedule/archives/empstat_nr.htm#current.

U.S. Bureau of Labor Statistics released a detailed report on the Employment Situation of Veterans-2012 on March 20, 2013 (annual release each March). The table below shows some key figures from this report. Here is a link to the report: <http://www.bls.gov/news.release/pdf/lvet.pdf>

Unemployment Rates, annual averages

	2012	2011
Total Veterans, 18 & older	7.0	8.3
Gulf War-era II Veterans	9.9	12.1
Nonveterans	7.9	8.7

Gulf War era II

	2012	2011
Total, 18 & Over	9.9	12.1
18 - 24	20.4	30.2
25 - 34	10.6	13.0
35 - 44	5.0	6.0
45 - 54	7.7	4.1
55 - 64	5.8	7.8

Please note, because the estimates for the veterans' annual average unemployment provided in the Veterans' Employment Situation - 2012 are more accurate than the other estimates discussed here, this is likely the best data available for users to reference in public presentations, grant applications, etc.

State Data

Employment and unemployment statistics by Veteran status for the state of Illinois are only available annually. These annual averages are produced by the U.S. Bureau of Labor Statistics for a special request from the Veteran's Employment and Training Services (VETS) in the Department of Labor. These figures are used by VETS in their State allocations under the Jobs for Veterans Act. The Illinois unemployment rate for Veterans aged 20 years and older in 2012 was 6.7%. The rate in 2011 was 8.3%.

Advanced Options

Download All Data

Clear All Selections

Multi-Select Mode

Help

Loading...

pdftables
tabula
messytables

Use the right types:

Telephone numbers are not numbers!

Parsable / canonical
formats make better indexes

Fix data before
you need it fixed

Data has a tendency
to be used in
unanticipated ways.

The background image shows a panoramic view of the city of Bogotá, Colombia, during sunset. The city is nestled in a valley, with its dense urban sprawl stretching towards the horizon. In the foreground, a steep, densely forested hillside covered in lush green trees and shrubs slopes down towards the city. The sky above is filled with dramatic, layered clouds, with the warm light of the setting sun casting a golden glow over the city's buildings and the surrounding landscape.

Thank You!

@makmanalp