

Welcome! The webinar will begin shortly.

DG CNECT

Directorate-General for Communications Networks, Content and Technology **DG DIGIT** Directorate-General for Digital Services



Some housekeeping

DIGITAL EUROPE PROGRAMME

- The Chat and Q&A widgets Bottom right of the screen
- Please post session-related questions in Q&A
- The session will be recorded and available shortly after the end of the session on our website

• Do

- Participate in the chat during the session
- Ask questions during the Q&A portion at the end
- Give feedback after the session
- Join the dedicated discussion board
- Don't
- Self-promote
- Disrespect anyone







Agenda

- Welcome [11:00-11:05]
- Session 4 Recap [11.05-11.15]
- $\,\circ\,$ Data Analytics [11:15-12:00]
- \circ Q&A [12:00-12:10]
- Wrap up [12:10 12:15]



What is the Big Data Test Infrastructure (BDTI)?

Not <u>only</u> for big data, for public sector in general (open data)

You have the key ingredients (datasets), we equip you with the best **open-source tool** to create amazing recipes for **public good**.

Six months free of charge service for EU public administrations * Ready-to-use data analytics stack and support Cloud platform based on open-source tools



* The cost of the pilot project must fit within the funding boundaries of the BDTI pilot budget

Who is the Big Data Test Infrastructure (BDTI) for?



European Public Administrations

All European Public Administrations at local, regional and national level can independently apply for a BDTI pilot project



Ecosystem with academia and private sector

Academia, spin-off, startups can apply for pilot projects once there is a **clear collaboration** with a Public Administration which will be the main point of contact for the project (**Master/PhD, GovTech startups**)



Are you working for a public administration in need of a free playground for data analytics?

https://big-data-test-infrastructure.ec.europa.eu/





Use case: "Identify Innovative Green Energy Projects from EU Universities"

Zoi and her team, work in the Educational Department of a highly polluted EU region. Their **purpose** is <mark>to identify potential</mark> "green energy partnerships" between EU universities to address pollution issues. Zoi reached out the BDTI team to learn how to **connect the dots**, use the free BDTI playground to **create** *insights* and produce a *report*, which can be **easily updated** when new data is available.





They found some interesting **open data**:

- List of innovation green energy projects across EU countries
- CO2 emissions of these countries



Recap: Zoi Data Journey

https://code.europa.eu/bdti/bdti-essentials-course



51



Recap Session 3 – Data Blending and Storage



What is Data Blending and Storage?



Blending	Export and Storage	
Stack new data at the bottom of a table → same columns, add rows	Export the data to store it in different file formats: CSV, Excel, Table, JSON.	
Join information from tables by defining matching criteria → add columns	Scalability , when things get bigger , and we have too much data, it is better to use Databases	

Data Access

Data Cleaning and Transformation Data Blending and Export

Analytics: aggregation, visualisation, reporting



EVOLUTION OF THE R&I FRAMEWORK PROGRAMMES BUDGET

The EU budget has been significantly rising over time from EUR 3 271 million in FP1 to EUR 96 899 million today for Horizon Europe and Euratom. This illustrates the clear ambition of the EU for its main R&I funding instrument and increases relevance of science and technology in society to address our challenges.





HORIZON



21

DIGITAL EUROPE PROGRAMME ⊜ ጭ № @ ₼

This budget compares favourably with the previous programme, Horizon 2020. In real terms, and excluding



Exercise Review



Summary: Data Blending and Storage



From multiple data files (8)

- Organisations (14-20/ 21-27)
- Projects (14-20/ 21-27)
- EuroSciVoc (14-20/ 21-27)
- CO2
- EU Country Names



To a **single** data file with all the information that we export and store







Session 4: Analytics, Visualisation, and Reporting



Learning Outcomes

At the end of this session, you should be able to:

- Sign into the BDTI framework, initialise KNIME, PgAdmin, PostreSQL, and Apache superset
- 2. Read data from a data base to KNIME and Apache Superset
- 3. Make sense of the data by applying basic aggregation
- 4. Create visualisations and a basic report







Open for Innovation

How do we get insights?

1) Aggregation

Create a summarised version of the data

2) Pivoting

Rearrange data to analyse it from different perspectives.

3) Visualisation

Transform data tables to communicate their results visually





Analytics: from objectives to insights

Objective	Insights
Identify countries with the most university involvement in green projects	Aggregate number of projects and plot data with a stacked bar chart
Identity the most active EU- based universities involved in green projects	Aggregate number of projects and plot data with a bar chart
Identify trends in started projects per country per year	Plot data with a line plot
Explore the relation between number of projects per university, energy project funding per university and a country's CO2 emission	Filter plot data with a parallel co- ordinates plot







Data Aggregation



Data Aggregation

- Create a summarised version of the data
- Aggregate by one or more categories
- Use one or more aggregation methods



Use Case

Zoi has reached the point where the data can be **aggregated** to get the necessary **insight** concerning the energy projects and CO2 emissions



`ommiccio

Data Aggregation: grouping data

- Group data based on common attributes or categories.
- Summarise data within each group to gain insights.

Category	# Ordered Items			
Clothing	2			Group
Home	3			Clothing
Clothing	1			Home
Clothing	5			Electronics
Electronics	7		Aggregated on Category	Aggregated on Category
Electronics	5		(group) by Sum (aggregation	(group) by Sum (aggregation
	Category Clothing Home Clothing Clothing Electronics	Category# Ordered ItemsClothing2Home3Clothing1Clothing5Electronics7Electronics5	Category# Ordered ItemsClothing2Home3Clothing1Clothing5Electronics7Electronics5	Category# Ordered ItemsClothing2Home3Clothing1Clothing5Electronics7Solution5Electronics5





Data Aggregation: grouping data

Product ID	Category	# Ordered Items
P 1	Clothing	2
P 2	Home	3
P 3	Clothing	1
P 4	Clothing	5
P 5	Electronics	7
P 6	Electronics	5

Group	Sum(# Ordered Items)
Clothing	8
Home	3
Electronics	12

Group - Aggregate



Aggregation by counting number of projects per Uni



#	RowID	projectID String	name String	activityTy	countryN	street String	ecMaxCo Number (dou	postCode	city String	startDat Local Date
1	Row	773606	PLIOSAUR EN	PRC	United Kingd	34 Sally Hill	50,000	BS20 7BH	PORTISHEAD	2017-05
2	Row	734032	NOVA INNOV	PRC	United Kingd	45 TIMBER B.	2,250,266	EH6 6QH	Edinburgh	2016-10
3	Row	745862	SKF GMBH	PRC	Germany	GUNNAR WE.	14,914,599.5	97421	Schweinfurt	2017-07
4	Row	745862	AGENCE INT	PRC	Belgium	QUAI SAINT	14,914,599.5	7500	Tournai	2017-07
5	Row	745862	SKF BV	PRC	Netherlands	MEIDOORNK	14,914,599.5	3992 AE	Houten	2017-07
6	Row	745862	RSK ENVIRO	PRC	France	202 QUAI DE .	14,914,599.5	92110	Clichy	2017-07
7	Row	745862	MOJO MARIT	PRC	United Kingd	27 AVENUE R.	14,914,599.5	TR11 4AY	Falmouth	2017-07
8	Row	745862	SKF MARINE	PRC	Germany	HERMANN B.	14,914,599.5	20457	HAMBURG	2017-0
9	Row	745862	WOOD GROU	PRC	France	60 RUE DE LA	14,914,599.5	75009	Paris	2017-0
10	Row	745862	OFFSHORE R	REC	United Kingd	OFFSHORE H.	14,914,599.5	NE24 1LZ	Blyth Northu	2017-0
11	Row	745862	NOVA INNOV	PRC	United Kingd	45 TIMBER B.	14,914,599.5	EH6 6QH	Edinburgh	2017-0
12	Row	745862	HMK TECHNI	PRC	United Kingd	KAPPA HOUS	14,914,599.5	CW12 1QJ	Congleton Ch	2017-0
13	Row	745862	THE UNIVER	HES	United Kingd	OLD COLLEG.	14,914,599.5	EH8 9YL	Edinburgh	2017-0
14	Row	745862	JAMES FISH	PRC	United Kingd	MICHAELSO	14,914,599.5	LA14 1HR	Barrow-In- Fu	2017-0
15	Row	745862	SKF (U.K) LI	PRC	United Kingd	SUNDON PAR	14,914,599.5	LU3 3BL	Luton	2017-0
16	Row	745855	OPENHYDRO	PRC	France	4 - 6 AVENUE	15,000,000	50106	Cherbourg-Oc	2017-0
17	Row	745855	CRIST OFFSH	PRC	Poland	UL. CZECHO	15,000,000	81 336	Gdynia	2017-0
18	Row	745855	THE EUROPE	отн	United Kingd	THE CHARLE.	15,000,000	KW16 3AW	Stromness	2017-0
19	Row	745855	DANMARKS	HES	Denmark	ANKER ENGE.	15,000,000	2800	Kongens Lyn	2017-0
20	Row	745855	OPENHYDRO	PRC	Ireland	1 CUSTOM H.	15,000,000	D01 C2CS	Dublin 1	2017-0
21	Row	745855	NAVAL ENER	PRC	France	31-35 RUE DE	15,000,000	75015	Paris	2017-0
22	Row	727793	SKF GMBH	PRC	Germany	GUNNAR WE	4,401,565.25	97421	Schweinfurt	2016-1
23	Row	727793	RHEINISCH	HES	Germany	TEMPLERGR.	4,401,565.25	52062	Aachen	2016-1
24	Row	727793	SKF MARINE	PRC	Germany	HERMANN B.	4,401,565.25	20457	HAMBURG	2016-1
25	Row	727793	SIEMENS PLC	PRC	United Kingd	SIR WILLIAM	4,401,565.25	GU16 8QD	Camberley	2016-1
26	Row	727793	WOOD GROU	PRC	France	60 RUE DE LA	4,401,565.25	75009	Paris	2016-1

Rows: 28 | Columns: 2 # RowID 1 2 3 4 5 6 7 8 9 101 11 12 13 14 15 16 18 Ro Ro Belgiun Ro Bulgaria Ro Croatia 14 Ro Cyprus 10 Ro Czech Republ Ro Denmar Ro Estonia Ro Finland 33 Ro France 1753 Ro Germany 2142 Ro Greece 545 Rov Hungary 14 Ro Ireland 251 1763 Italy Ro Latvia 72 Row Lithuani

Group - Aggregate



Aggregate in the different BDTI tools



SQL		R	
SELECT SUM(column_to_aggregate) FROM table1 GROUP BY column_to_group;df <- read.cs aggregate(c by=list(Cate FUN=sum)		sv('table1.csv') If\$column_to_ag gory=df\$column_	gregate, _to_group),
PYTHON import pandas as pd df = pd.read_csv('table1.csv') grouped_df = df.groupby(['column_to_group']).sum()		KNIME Row Aggregator	GroupBy



Pivoting

- Rearrange data to analyse it from different perspectives.
- Transform rows into columns or vice versa to facilitate analysis.

Product ID	Store	Category	# Ordered Items	
P 1	Online	Clothing	2	
P 2	Onsite	Home	3	
P 3	Onsite	Clothing	1	\langle
P 4	Online	Clothing	5	
P 5	Online	Electronics	7	
P 6	Online	Electronics	5	

Aggregation: Count

Category	Online	Onsite
Clothing	2	1
Home	0	1
Electronics	2	0

Aggregation: Sum (# Ordered Items)

Category	Online	Onsite
Clothing	7	1
Home	0	3
Electronics	12	0



Pivoting

Product ID	Store	Category	# Ordered Items
P 1	Online	Clothing	2
P 2	Onsite	Home	3
P 3	Onsite	Clothing	1
P 4	Online	Clothing	5
P 5	Online	Electronics	7
P 6	Online	Electronics	5

Aggregation: Sum (# Ordered Items)

	Category	Online	Onsite
	Clothing	7	1
	Home	0	3
	Electronics	12	0

Group - Pivot - Aggregate



Pivot to show number of projects by project type



Rows	5: 21476	Colu	imns: 17								
	#	RowID	projectID String	name String	activityTy	countryN String	street String	ecMaxCo Number (dou	postCode String	city _{String}	startDat
	1	Row	773606	PLIOSAUR EN	PRC	United Kingd	34 Sally Hill	50,000	BS20 7BH	PORTISHEAD	2017-05
	2	Row	734032	NOVA INNOV	PRC	United Kingd	45 TIMBER B.	2,250,266	EH6 6QH	Edinburgh	2016-10
	3	Row	745862	SKF GMBH	PRC	Germany	GUNNAR WE	14,914,599.5	97421	Schweinfurt	2017-07
	4	Row	745862	AGENCE INT	PRC	Belgium	QUAI SAINT	14,914,599.5	7500	Tournai	2017-07
	5	Row	745862	SKF BV	PRC	Netherlands	MEIDOORNK	14,914,599.5	3992 AE	Houten	2017-07
	6	Row	745862	RSK ENVIRO	PRC	France	202 QUAI DE .	14,914,599.5	92110	Clichy	2017-07
	7	Row	745862	MOJO MARIT	PRC	United Kingd	27 AVENUE R.	14,914,599.5	TR11 4AY	Falmouth	2017-07
	8	Row	745862	SKF MARINE	PRC	Germany	HERMANN B.	14,914,599.5	20457	HAMBURG	2017-07
	9	Row	745862	WOOD GROU	PRC	France	60 RUE DE LA	14,914,599.5	75009	Paris	2017-07
	10	Row	745862	OFFSHORE R	REC	United Kingd	OFFSHORE H.	14,914,599.5	NE24 1LZ	Blyth Northu	2017-07
	11	Row	745862	NOVA INNOV	PRC	United Kingd	45 TIMBER B.	14,914,599.5	EH6 6QH	Edinburgh	2017-07
	12	Row	745862	HMK TECHNI	PRC	United Kingd	KAPPA HOUS	14,914,599.5	CW12 1QJ	Congleton Ch	2017-07
	13	Row	745862	THE UNIVER	HES	United Kingd	OLD COLLEG.	14,914,599.5	EH8 9YL	Edinburgh	2017-07
	14	Row	745862	JAMES FISH	PRC	United Kingd	MICHAELSO	14,914,599.5	LA14 1HR	Barrow-In- Fu	2017-07
	15	Row	745862	SKF (U.K) LI	PRC	United Kingd	SUNDON PAR	14,914,599.5	LU3 3BL	Luton	2017-07
	16	Row	745855	OPENHYDRO	PRC	France	4 - 6 AVENUE	15,000,000	50106	Cherbourg-Oc	2017-09
	17	Row	745855	CRIST OFFSH	PRC	Poland	UL. CZECHO	15,000,000	81 336	Gdynia	2017-09
	18	Row	745855	THE EUROPE	отн	United Kingd	THE CHARLE.	15,000,000	KW16 3AW	Stromness	2017-09
	19	Row	745855	DANMARKS	HES	Denmark	ANKER ENGE.	15,000,000	2800	Kongens Lyn	2017-09
	20	Row	745855	OPENHYDRO	PRC	Ireland	1 CUSTOM H.	15,000,000	D01 C2CS	Dublin 1	2017-09
	21	Row	745855	NAVAL ENER	PRC	France	31-35 RUE DE	15,000,000	75015	Paris	2017-09
	22	Row	727793	SKF GMBH	PRC	Germany	GUNNAR WE	4,401,565.25	97421	Schweinfurt	2016-11
	23	Row	727793	RHEINISCH	HES	Germany	TEMPLERGR	4,401,565.25	52062	Aachen	2016-11
	24	Row	727793	SKF MARINE	PRC	Germany	HERMANN B.	4,401,565.25	20457	HAMBURG	2016-11
	25	Row	727793	SIEMENS PLC	PRC	United Kingd	SIR WILLIAM	4,401,565.25	GU16 8QD	Camberley	2016-11
	26	Row	727793	WOOD GROU	PRC	France	60 RUE DE LA	4,401,565.25	75009	Paris	2016-11

] #	Rowl	countryName ↓ String	HES Number (integer)	\sim	OTH Number (integer)	\sim	PRC Number (integer)	\sim	PUB Number (integer)	\sim	REC Number (integer)
1	Row	Austria	127								
2	Row	Belgium	15		334		514		59		245
3	Row	Bulgaria	_0		27		76		17		40
4	Rowa	Croatia	38		20		73		25		28
5	Row4	Cyprus	39		11		79		15		4
6	Rows	Czech Republic	86		39		145		22		58
7	Rowe	Denmark	221		60		229		34		23
8	Row	Estonia	26		25		54		15		5
9	Rowa	Finland	124		19		206		19		124
10	Rows	France	351		122		1323		74		542
11	Row.	Germany	626		148		1328		35		715
12	Row.	Greece	175		48		392		34		214
13	Row.	Hungary	26		24		86		17		27
14	Row.	Ireland	97		24		160		24		9
15	Row.	Italy	561		138		1285		88		408
16	Row.	Latvia	16		20		20		26		8
17	Row.	Lithuania	13		7		35		7		18
18	Row.	Luxembourg	7		3		52		6		12
19	Row.	Malta	5		5		12		8		1
20	Row.	Netherlands	326		89		810		53		237
21	Row.	Poland	70		37		153		32		77
22	Row.	Portugal	66		47		192		21		117
23	Row.	Romania	43		45		80		32		26
24	Row.	Slovakia	21		9		51		14		7
25	Row.	Slovenia	25		16		108		21		83
26	Row.	Spain	357		136		1265		165		802
27	Row.	Sweden	18		24		297		33		84
28	Row.	United Kingdom	709		100		720		48		52

Group - Pivot - Aggregate



Pivot in the different BDTI tools



European Commission

SQL	R
SELECT "Category", SUM ("# Ordered Items") filter (WHERE "Store" = "Onsite") AS 'Onsite', SUM ("# Ordered Items") filter (WHERE "Store" = "Online") AS	library(pivottabler) df <- read.csv('table1.csv') qhpvt (df, "Store", "Category", " sum (# Ordered Items)")
'Online' FROM #table# GROUP BY "Category";	KNIME
PYTHON	
<pre>import pandas as pd df = pd.read_csv('table1.csv') pivoted_df = pd.pivot_table(df, values=['Store'], index=["Category"], aggfunc={'# Ordered Items': "sum"})</pre>	

Data Aggregation - Summary

- Summarised version of the data
- Aggregate by one or more categories
- **Pivot** for a more detailed aggregation







Visualisation



How do we get insights?

1) Aggregation Create a summarised version of the data

2) Pivoting Rearrange data to analyse it from different perspectives.

3) Visualisation

Transform data tables to communicate their results visually



Data Visualisation

- Transform data tables to communicate their results visually
- Help data understanding and get more insights
- Enable identifying patterns are trends that are hidden in tabular format

Use Case

To see the insights in the data not visible in a table Zoe's team is looking for some helpful visualisations

Analytics: from objectives to insights

Objective	Insights
Identify countries with the most university involvement in green projects	Aggregate number of projects and plot data with a stacked bar chart
Identity the most active EU- based universities involved in green projects	Aggregate number of projects and plot data with a bar chart
Identify trends in started projects per country per year	Plot data with a line plot
Explore the relation between number of projects per university, energy project funding per university and a country's CO2 emission	Filter plot data with a parallel co- ordinates plot

Commission

Visualisation with the BDTI tools

University name OCCURRENCE_COUNT

- KNIME 1)
- 2) Apache Superset

Project count

HES COTH PRC PUB REC

3) Metabase

Trends: New started projects per year per country

Relation between CO2, uni projects and total project costs

Now you should be able to:

- 1. Sign into the BDTI framework, initialise KNIME, PgAdmin, PostreSQL, and Apache superset
- 2. Read data from a data base to KNIME and Apache Superset
- 3. Make sense of the data by applying basic aggregation
- 4. Create visualisations and a basic report

Data Access Data Cleaning and Transformation Data Blending and Storage Analytics: aggregation, visualisation, reporting Analytics

Next session (next week)

European Commission

Practice, practice and practice!

 Slides, hands-on exercises, and solutions will be provided on the webinar Gitlab repository

https://code.europa.eu/bdti/bd ti-essentials-course

BDTI Essentials Course Project ID: 635 Request Access		
-> 12 Commits 🖇 1 Branch 🧷 0 Tags 🗔 2 MiB Project Stor	age	
Upload banner session 1 Maria Claudia BODINO authored 4 days ago		9cceld4d
master \checkmark bdti-essentials-course / [+ \checkmark		History Find file Edit v Clone v
문 README 한 BSD 3-Clause "New" or "Revised" License	Auto DevOps enabled	
Name	Last commit	Last update
🗅 Session 1: Data Access and Exploration	Upload banner session 1	4 days ago
BDTI_Banner_generic.png	Upload New File	5 days ago
	Add LICENSE	5 days ago
M# README.md	Update README.md	5 days ago
B README.md		
Copyright 2023 European Union Documentation in this repository is licensed under the Crea	tive Commons Attribution 4.0 License, and code samples are lice	ensed under the BSD 3-Clause licence.
Eropan Commission		
From hype	to action:	IN EUROPE EUROPE PROGRAMME ▶ ■ 零 验 @ 最

European Commissior

Practice, practice and practice!

 Slides, hands-on exercises, and solutions will be provided on the webinar Gitlab repository

https://code.europa.eu/bdti/bd ti-essentials-course

Session 4 - Exercise

Analytics: Data Aggregation, Visualisation and Reporting

Zoi has the data ready to work, now she wants o build a dashboard with some visualisations and generate a pdf report from it.

Access data from a DB	Pivot countries	Visualizations
	Groupby and Rank CO2 emissions	

Your BDTI journey starts here

Congrats! You are on the first step to data-driven innovation.

Brainstorm your data project

Apply for BDTI Pilotlight process

Course discussion board

Welcome!

Here, you can ask questions and discuss topics related to the BDTI Essentials Online Course. The board is moderated by the BDTI team, so your questions will be answered quickly. We also encourage members to discuss and help each other where possible to build a collaborative space and rewarding community.

European Commission

Helpful links

Get in touch and follow the BDTI activities

Are you working for a public administration in need of infrastructure for data analytics?

Visit BDTI's website

Subscribe to BDTI's newsletter

Subscribe to BDTI's Joinup

https://big-data-test-infrastructure.ec.europa.eu/