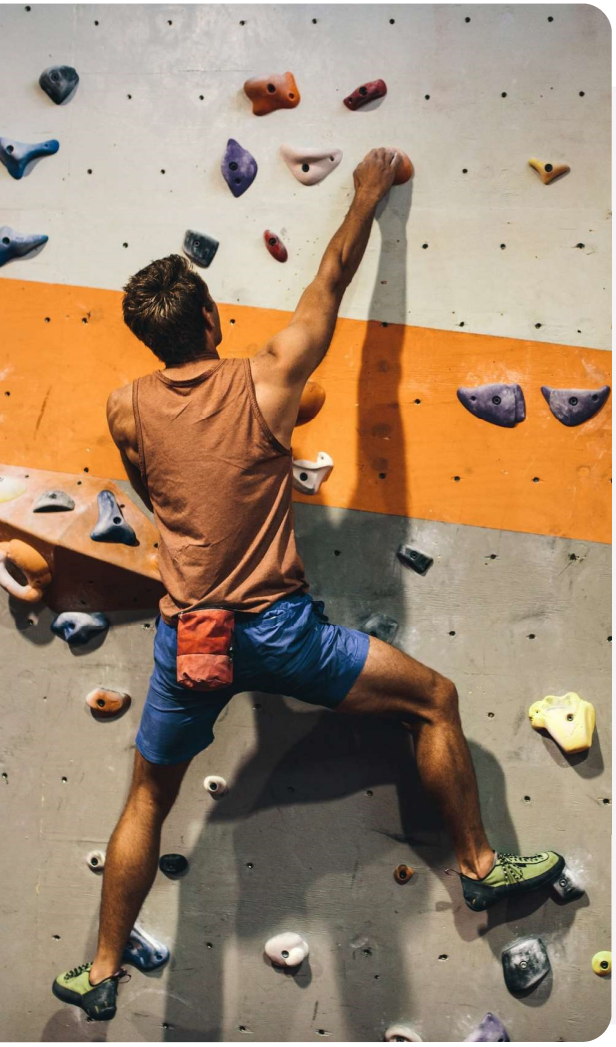
DATA ANALYST – BUSINESS INTELLIGENCE / EDITION 2024



SELECT THE RIGHT CANDIDATES FOR SPONSORSHIP IN SPORT CLIMBING

A BUSINESS INTELLIGENCE PROJECT

Stéphane JULLIEN
Bruxelles, DigitalCity, Sept 6th 2024

github.com/SJullienGithub/Climbing

BRUXELLES FORMATION
former pour l'emploi

Digitalcity.brussels

About Stephane JULLIEN

20 YEARS OF PROVIDING TECHNICAL SUPPORT TO INTERNAL AND EXTERNAL CUSTOMERS
IN MULTIPLE PROCESSES (MASTER DATA MNGT IN PLM/PDM AND SAP, WORK INSTRUCTION MNGT, BIOCHEMICAL PURIFICATION)
AND INDUSTRIES (BIO-PHARMACEUTICAL, RAILWAY, TRACTORS, AERONAUTICS

STRENGHTS

2002 : MASTER DEGREE IN BIOTECHNOLOGIES

2012 : LEAN MANUFACTURING / METHODES INDUSTRIELLES

2017 - 23 : MASTER DATA MANAGEMENT IN SAP AND PLM

- SAP MIGRATION 60K MATERIALS & 17K MBOMS
- INTERFACE PLM & ORCHESTRA TO SAP
- END-USER SUPPORT IN INDUSTRIAL DATA
- END USER TRAINING

WEAKNESSES

- ONLY EXCEL AS DATA ANALYSIS AND VISUALISATION TOOL
- EXCEL TOO SLOW FOR SAP MASTER DATA MIGRATION
- NO ACCESS TO IT TOOL DATABASES
- IT TOOLS AND PROCESS “TOO” ALSTOM SPECIFIC OR OUTDATED : SAP R3, ...
- LACK OF SKILLS FOR LARGE SCALE DATA ANALYSIS

THREATS

END OF 2023 : NO MATCH BETWEEN CV AND JOB REQUIREMENTS IN DATA ANALYSIS AND SAP

- POWER-BI, TABLEAU, SQL, SAP4HANA, SAP PLM, ...

IMPACT OF LLMS (CHATGPT) AND AI ON JOB MARKET

OPPORTUNITIES

DISCOVERED POWER-BI @ SONACA

EAGER TO LEARN NEW TOOLS :

- DATABASE, AI, LLMS, SAP4HANA, BIG DATA,...

EAGER TO BE MORE EFFICIENT & AUTONOMOUS WITH DATA

DATA ANALYST COURSE @ DIGITAL CITY ... TO START

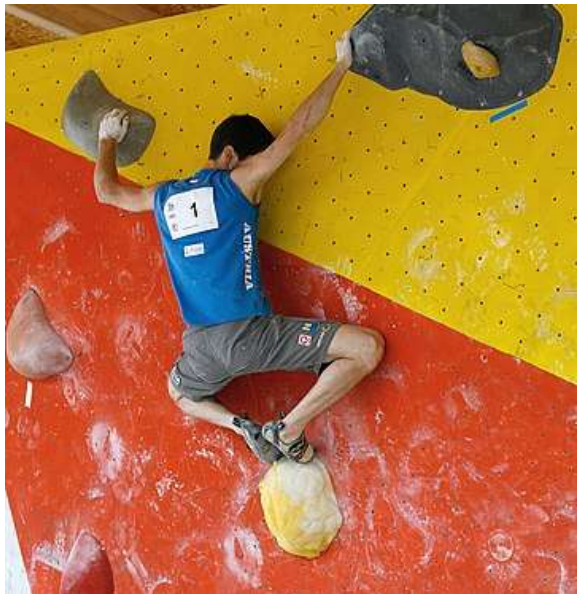
About Competition Climbing

LEAD

Competitors have 6 minutes to climb a 15-metre pre-bolted sport climbing route. They must also clip their safety rope into the various quickdraws while they ascend the route; failing to clip into a quickdraw terminates their climb at that position

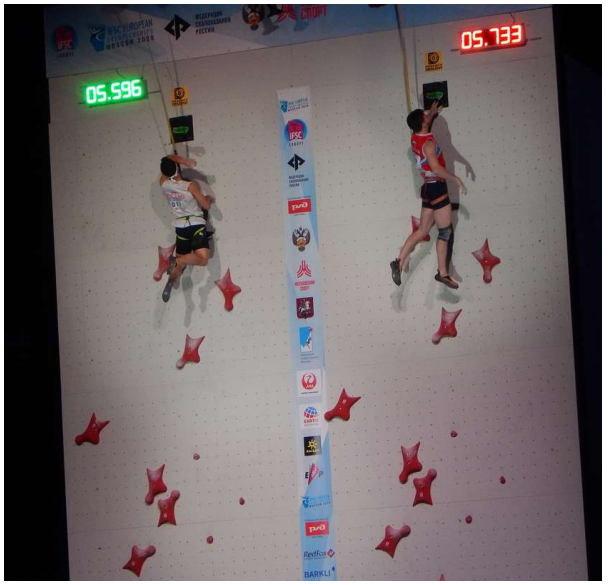
https://en.wikipedia.org/wiki/Competition_climbing

BOULDERING



Competitors have to "solve" multiple short 4.5-metre bouldering problems over a set time period, with the fewest falls. More complex / lead climbing. Competitors do not use a rope or any climbing protection, but crash pads that are laid across the ground for safety

SPEED



Competitors must race against the clock a 15-metre, standardised climbing wall, where the holds are always the exact same size and placed in the exact same location. As the emphasis is on speed, the climbers use an auto-belay top rope for protection.

COMBINED

About Competition Climbing

DISCIPLINES : Lead Bouldering Speed Combined

COMPETITIONS: Worldwide only
Summer Olympics IFSC World Cup
IFSC World Championship

RANK: Gold Silver Bronze

AGE CATEGORIES: years old
Youth_B 14 - 15 yo
Youth_A 16 - 17
Junior 18 - 19
Senior 20 and over

GENDER

BI Project Context

- A worldwide company specialized in Climbing Furnitures
- Has developed a connected climbing harness
- The marketing team suggests sponsorship of some competition climbers :
  - 2 World Senior Competition Climber (1 M / 1 F)
  - 2 Future Stars (1 M / 1 F) < 18 years in 2024:
  - A National Team
- The Top Management requests his newly Data Analyst to provide Top 3 Names for each
- Data Analyst’s solution : create a database gathering Climbers performance ranking in worldwide competitions, and over the years. Idem for Nations.

Sources, Process and Technologies

Search

Google Dataset Search,
Kaggle, ChatGPT, Wikipedia

- Lot of Climbing content in webpages,
- 1 dataset (2 csv > 9000 lines) on kaggle
- Data on Nations (PIB, Age, Altitude, ...) in Wikipedia

Ask

Google Form (draft),
Mails to Bxl climbing gyms,
Mail for Requesting API key

- Time spending
- No answer

Extract
Clean

- Webscrapping : Python Pandas + BeautifulSoup
- Pandas df. : delete multiindexing, column rename, type changes, slicing, ...

Wikipedia : html tables (compatible)

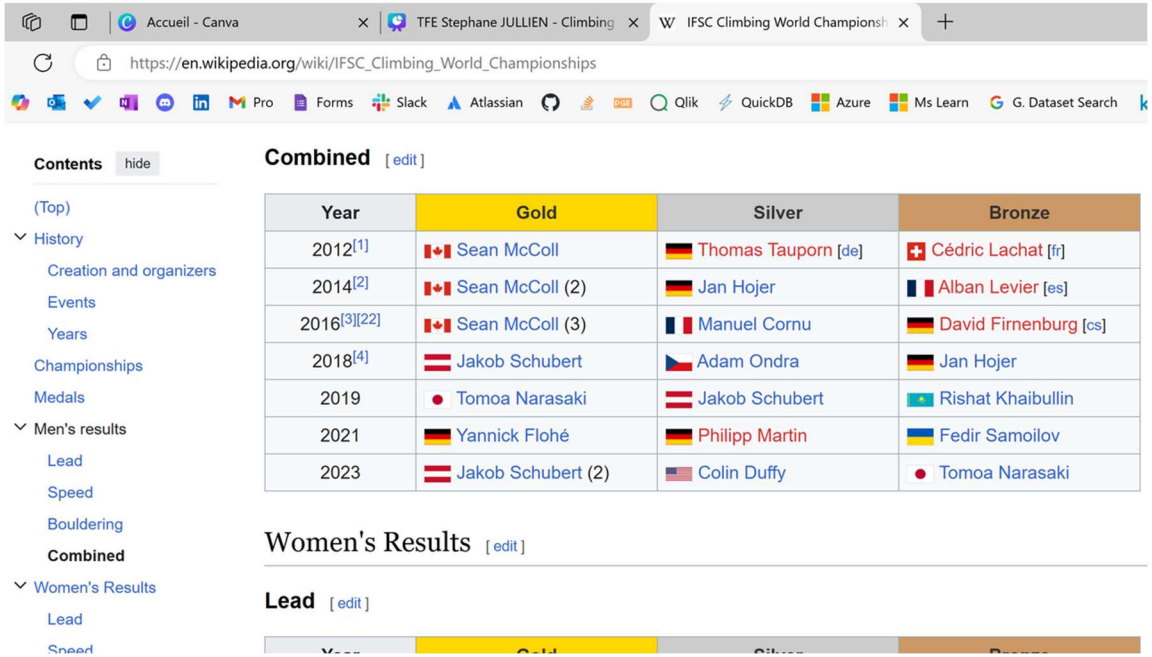
- ~ 30 tables / webpage :
- no standard structure > no automatization
- legends in data frame, comma missing,
- Competitor Nation displayed in icon in html and not webscrapped :
- languages of Nation names
- decimal comma disappearing

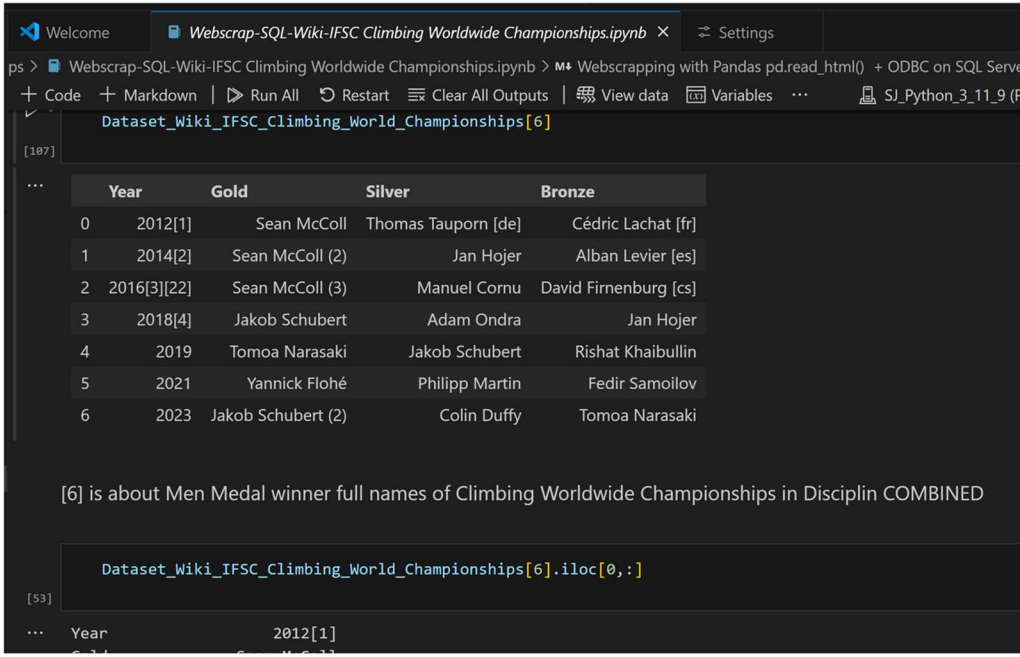
Other websites:

- no html tables > Selenium, API Call,
- Or no access (firewall, ...)

PowerQuery on webpages ? results

Sources, Process and Technologies





Load in DB Staging

- Python connected to SQL server : pyodbc for connecting and populate tables in SQL server
- DBMS SQL Server SSMS

Merge Tables with Python before loading in DB Staging

Sources, Process and Technologies

ETL within DB Staging

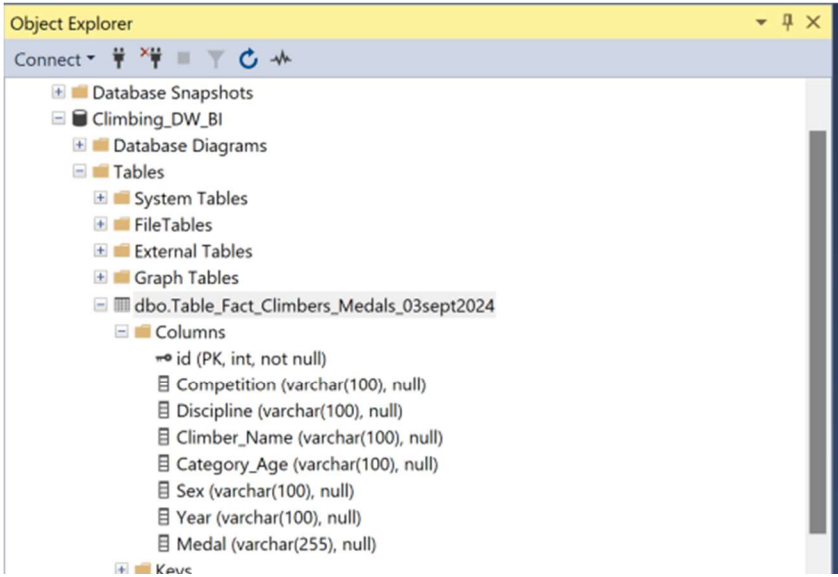
- SSIS to clean data and gather competitor tables to 4 tables (1 / competition), add extra columns

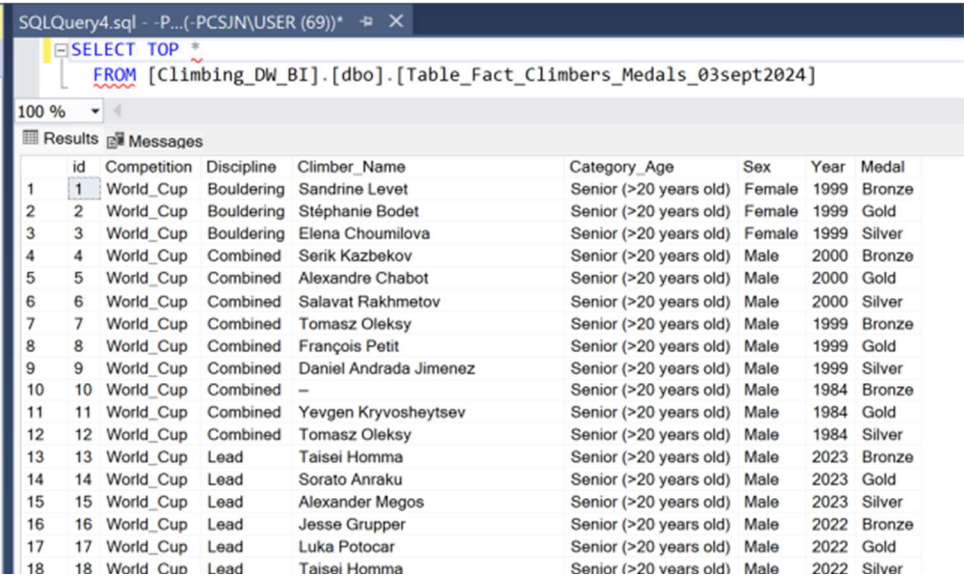
Cleansing : done in Python and tables reloaded into SQL Server
'Year' Type (Int or Str), Extra columns- in Unicode

ETL to DW

- SSIS : Unpivot Climber - Medal

> 1st Fact Table : "Competitors - Medals" (2088 records)
> 2nd Fact Table : "Nations - Medals "
> 4 Dim tables : GDP/inhabitant, Median Age, % by Category of Age, Altitude, ...)





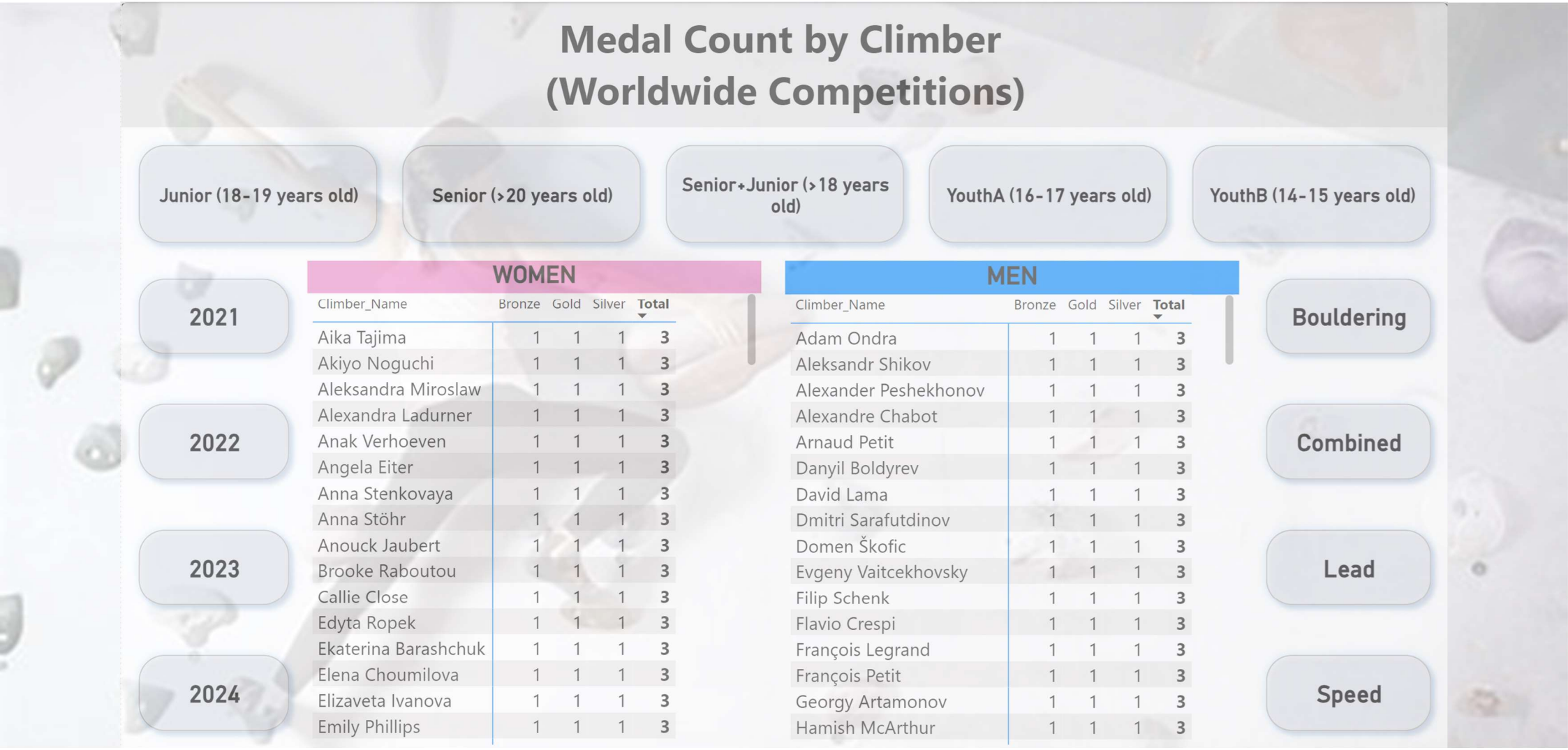
	id	Competition	Discipline	Climber_Name	Category_Age	Sex	Year	Medal
1	1	World_Cup	Bouldering	Sandrine Levet	Senior (>20 years old)	Female	1999	Bronze
2	2	World_Cup	Bouldering	Stéphanie Bodet	Senior (>20 years old)	Female	1999	Gold
3	3	World_Cup	Bouldering	Elena Choumilova	Senior (>20 years old)	Female	1999	Silver
4	4	World_Cup	Combined	Serik Kazbekov	Senior (>20 years old)	Male	2000	Bronze
5	5	World_Cup	Combined	Alexandre Chabot	Senior (>20 years old)	Male	2000	Gold
6	6	World_Cup	Combined	Salavat Rakhmetov	Senior (>20 years old)	Male	2000	Silver
7	7	World_Cup	Combined	Tomasz Oleksy	Senior (>20 years old)	Male	1999	Bronze
8	8	World_Cup	Combined	François Petit	Senior (>20 years old)	Male	1999	Gold
9	9	World_Cup	Combined	Daniel Andrada Jimenez	Senior (>20 years old)	Male	1999	Silver
10	10	World_Cup	Combined	–	Senior (>20 years old)	Male	1984	Bronze
11	11	World_Cup	Combined	Yevgen Kryvosheytsev	Senior (>20 years old)	Male	1984	Gold
12	12	World_Cup	Combined	Tomasz Oleksy	Senior (>20 years old)	Male	1984	Silver
13	13	World_Cup	Lead	Taisei Homma	Senior (>20 years old)	Male	2023	Bronze
14	14	World_Cup	Lead	Sorato Anraku	Senior (>20 years old)	Male	2023	Gold
15	15	World_Cup	Lead	Alexander Megos	Senior (>20 years old)	Male	2023	Silver
16	16	World_Cup	Lead	Jesse Grupper	Senior (>20 years old)	Male	2022	Bronze
17	17	World_Cup	Lead	Luka Potocar	Senior (>20 years old)	Male	2022	Gold
18	18	World_Cup	Lead	Taisei Homma	Senior (>20 years old)	Male	2022	Silver

Sources, Process and Technologies

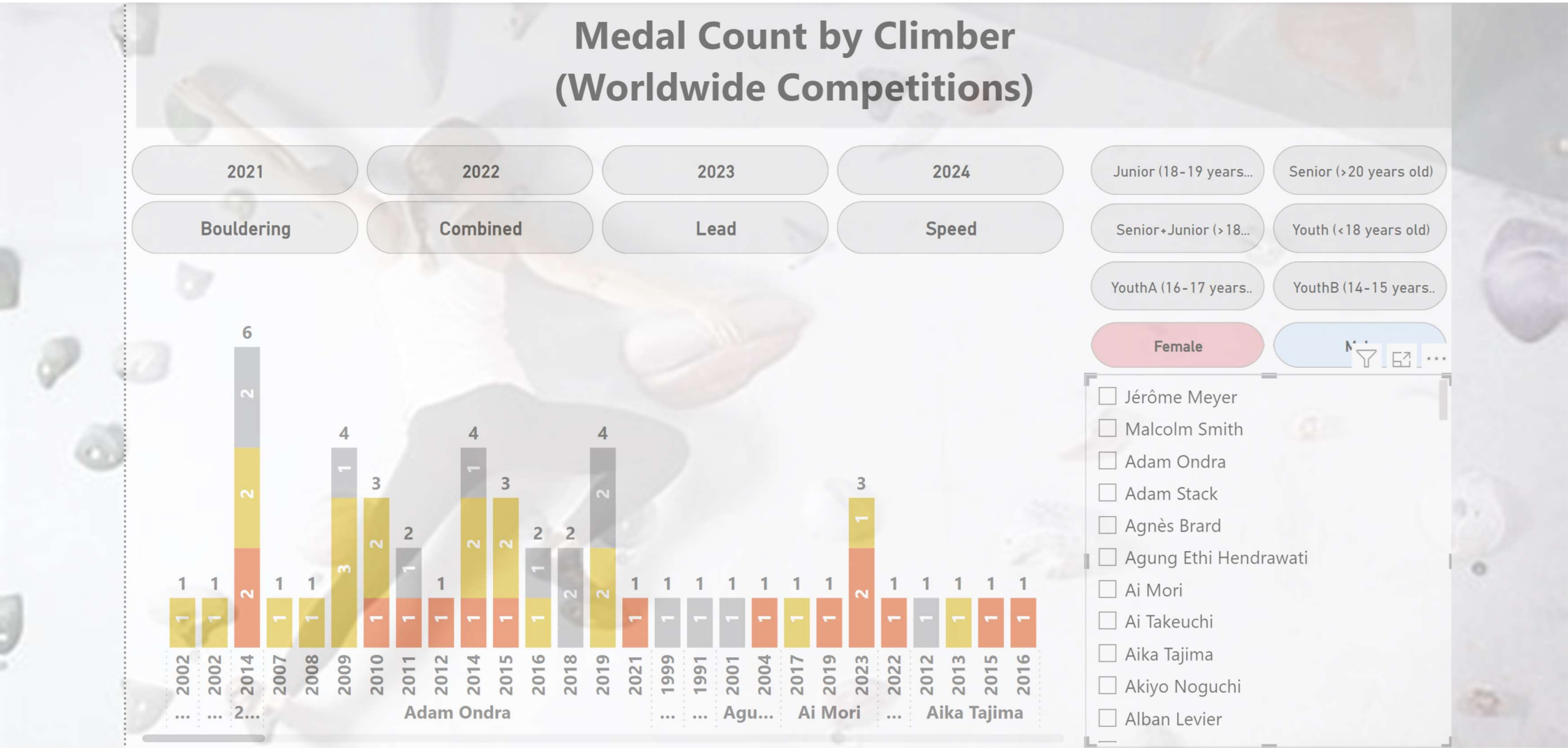
Data Viz

- PowerBI from DW in SQL Server
- Data Transformation

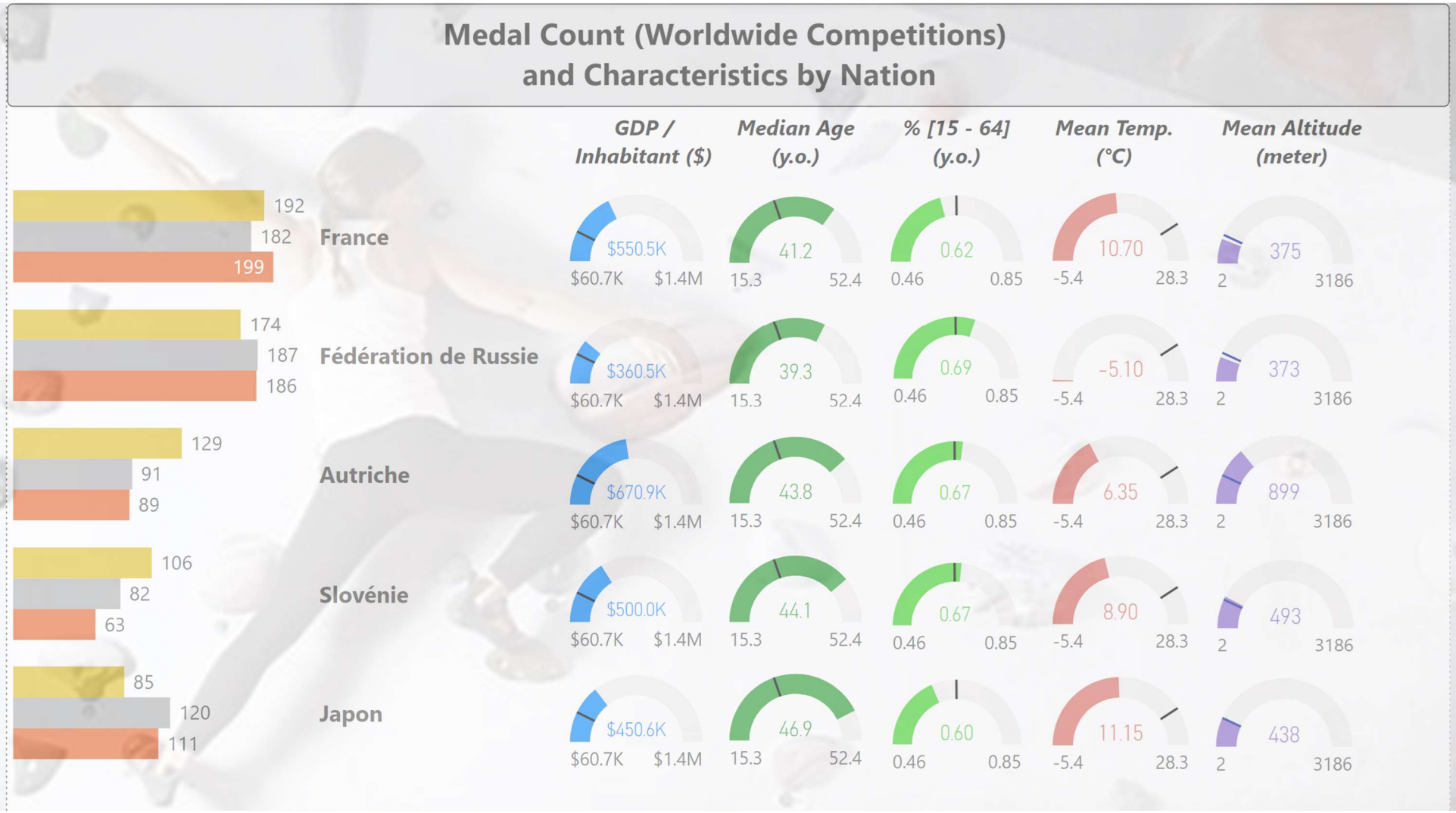
Business : Which climbers sponsorised ?



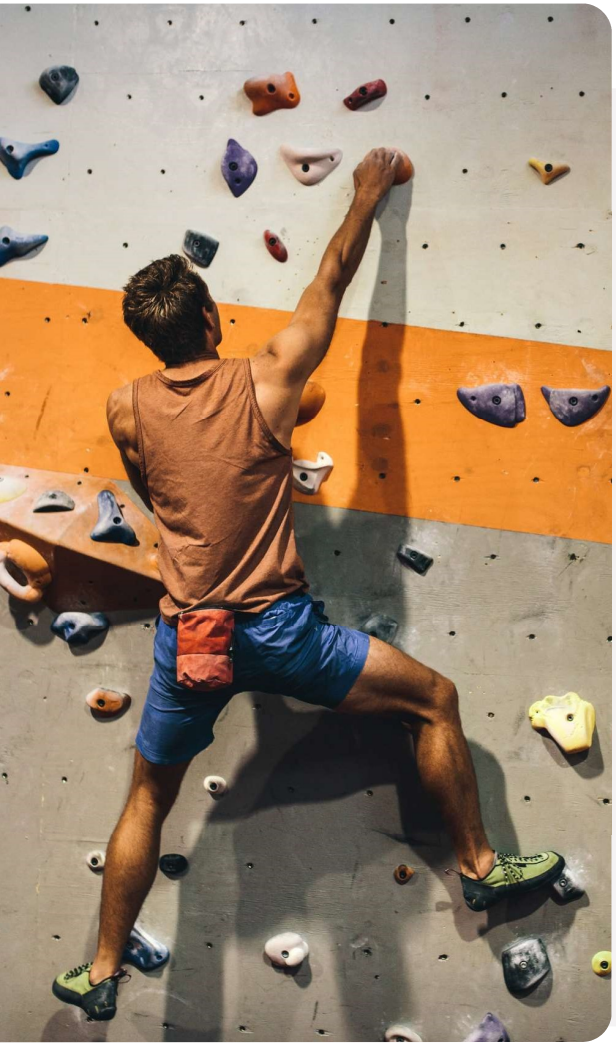
Business : Which climbers sponsorised ?



Business : Which Nation sponsorised ?



DATA ANALYST – BUSINESS INTELLIGENCE / EDITION 2024

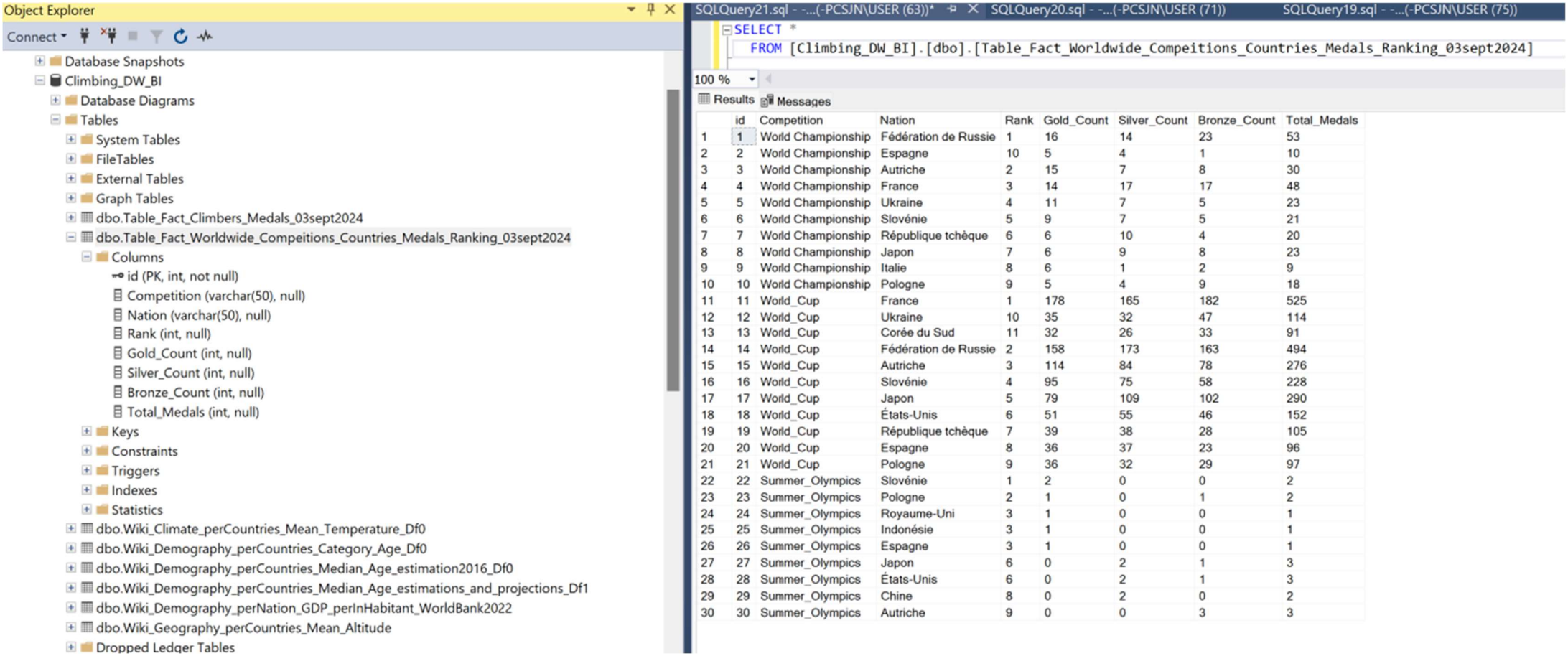


CONCLUSION
&
THANK YOU

Stéphane JULLIEN

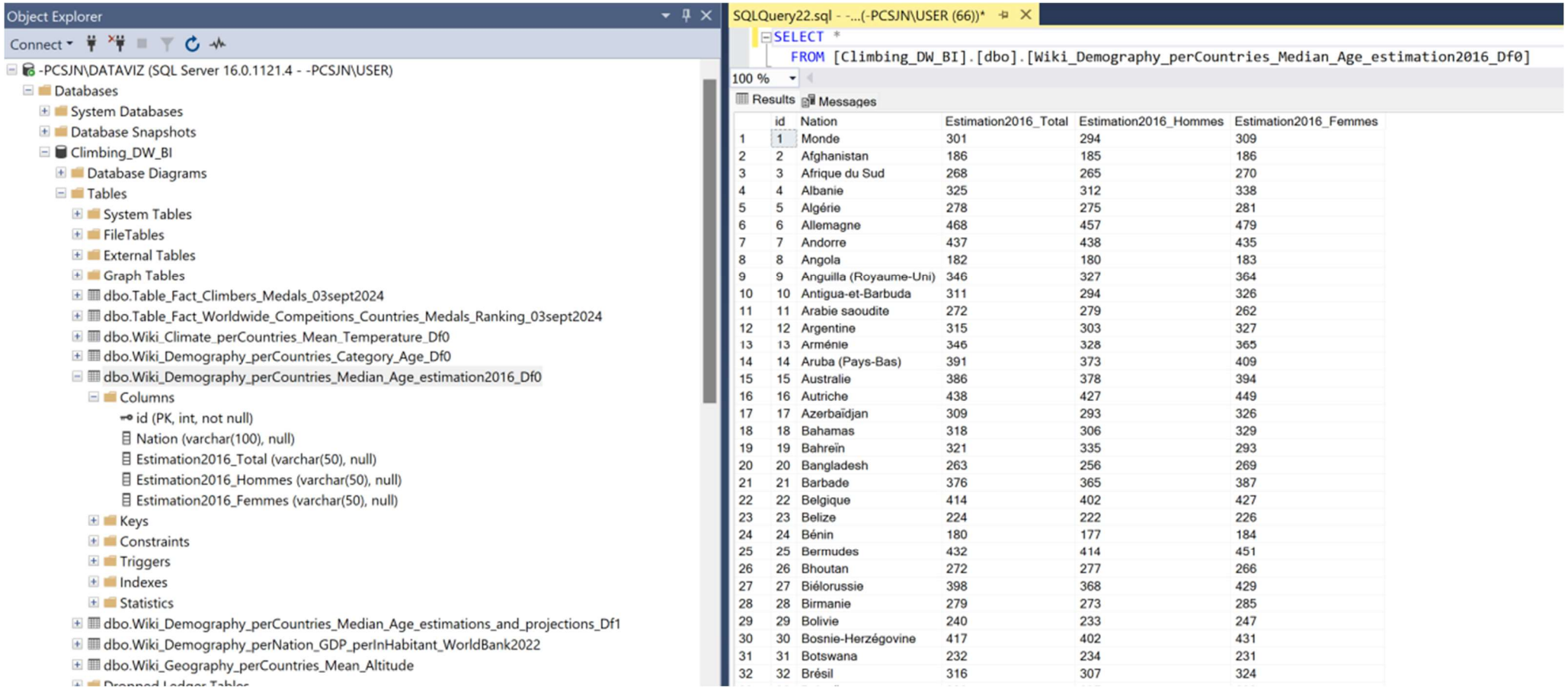
Bruxelles DigitalCity, Sept 6th 2024

Appendix



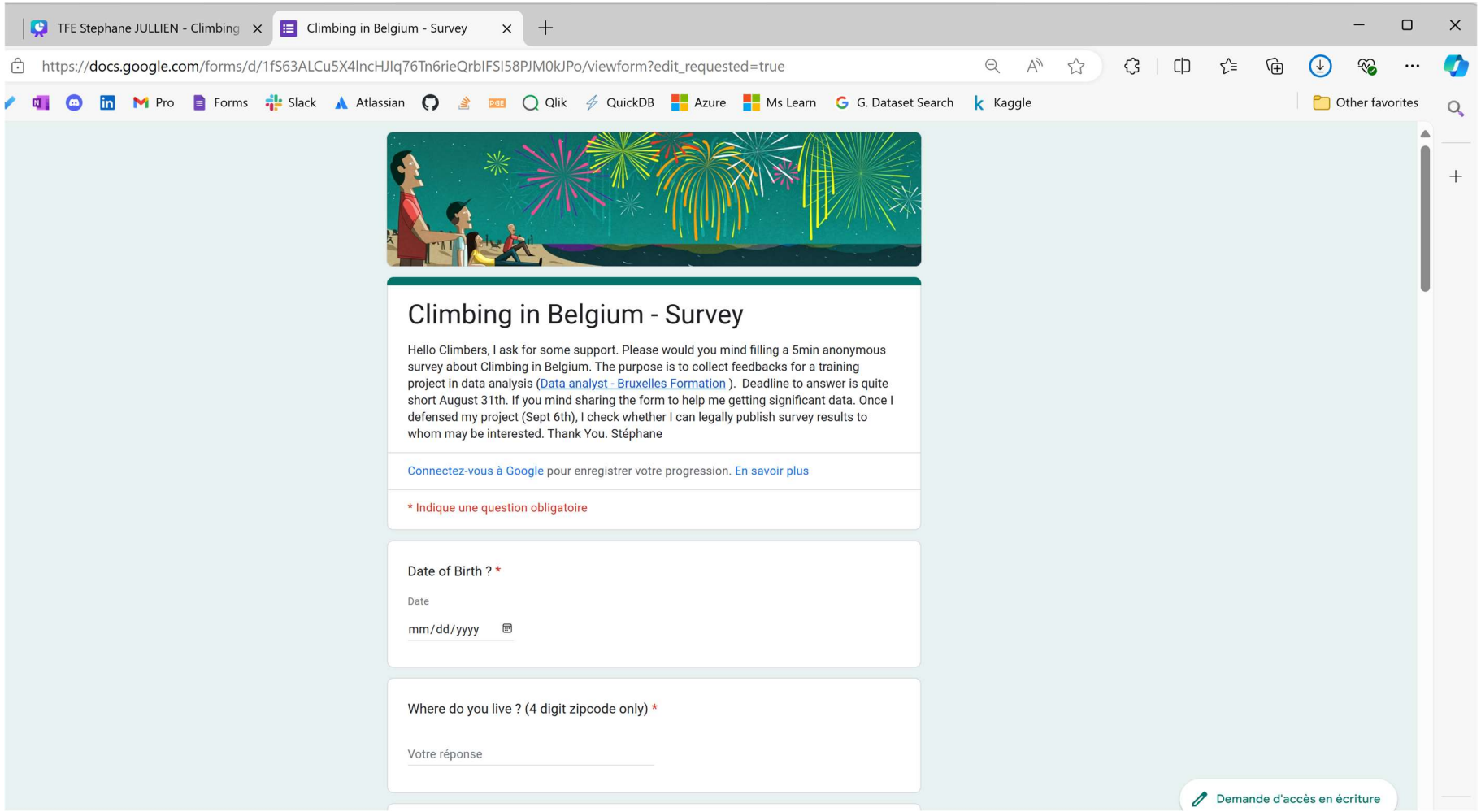
	id	Competition	Nation	Rank	Gold_Count	Silver_Count	Bronze_Count	Total_Medals
1	1	World Championship	Fédération de Russie	1	16	14	23	53
2	2	World Championship	Espagne	10	5	4	1	10
3	3	World Championship	Autriche	2	15	7	8	30
4	4	World Championship	France	3	14	17	17	48
5	5	World Championship	Ukraine	4	11	7	5	23
6	6	World Championship	Slovénie	5	9	7	5	21
7	7	World Championship	République tchèque	6	6	10	4	20
8	8	World Championship	Japon	7	6	9	8	23
9	9	World Championship	Italie	8	6	1	2	9
10	10	World Championship	Pologne	9	5	4	9	18
11	11	World_Cup	France	1	178	165	182	525
12	12	World_Cup	Ukraine	10	35	32	47	114
13	13	World_Cup	Corée du Sud	11	32	26	33	91
14	14	World_Cup	Fédération de Russie	2	158	173	163	494
15	15	World_Cup	Autriche	3	114	84	78	276
16	16	World_Cup	Slovénie	4	95	75	58	228
17	17	World_Cup	Japon	5	79	109	102	290
18	18	World_Cup	États-Unis	6	51	55	46	152
19	19	World_Cup	République tchèque	7	39	38	28	105
20	20	World_Cup	Espagne	8	36	37	23	96
21	21	World_Cup	Pologne	9	36	32	29	97
22	22	Summer_Olympics	Slovénie	1	2	0	0	2
23	23	Summer_Olympics	Pologne	2	1	0	1	2
24	24	Summer_Olympics	Royaume-Uni	3	1	0	0	1
25	25	Summer_Olympics	Indonésie	3	1	0	0	1
26	26	Summer_Olympics	Espagne	3	1	0	0	1
27	27	Summer_Olympics	Japon	6	0	2	1	3
28	28	Summer_Olympics	États-Unis	6	0	2	1	3
29	29	Summer_Olympics	Chine	8	0	2	0	2
30	30	Summer_Olympics	Autriche	9	0	0	3	3

Appendix



	id	Nation	Estimation2016_Total	Estimation2016_Hommes	Estimation2016_Femmes
1	1	Monde	301	294	309
2	2	Afghanistan	186	185	186
3	3	Afrique du Sud	268	265	270
4	4	Albanie	325	312	338
5	5	Algérie	278	275	281
6	6	Allemagne	468	457	479
7	7	Andorre	437	438	435
8	8	Angola	182	180	183
9	9	Anguilla (Royaume-Uni)	346	327	364
10	10	Antigua-et-Barbuda	311	294	326
11	11	Arabie saoudite	272	279	262
12	12	Argentine	315	303	327
13	13	Arménie	346	328	365
14	14	Aruba (Pays-Bas)	391	373	409
15	15	Australie	386	378	394
16	16	Autriche	438	427	449
17	17	Azerbaïdjan	309	293	326
18	18	Bahamas	318	306	329
19	19	Bahreïn	321	335	293
20	20	Bangladesh	263	256	269
21	21	Barbade	376	365	387
22	22	Belgique	414	402	427
23	23	Belize	224	222	226
24	24	Bénin	180	177	184
25	25	Bermudes	432	414	451
26	26	Bhoutan	272	277	266
27	27	Biélorussie	398	368	429
28	28	Birmanie	279	273	285
29	29	Bolivie	240	233	247
30	30	Bosnie-Herzégovine	417	402	431
31	31	Botswana	232	234	231
32	32	Brésil	316	307	324

Appendix


TFE Stephane JULLIEN - Climbing
Climbing in Belgium - Survey
https://docs.google.com/forms/d/1fS63ALCu5X4lncHJlq76Tn6rieQrbIFSI58PJM0kJPo/viewform?edit_requested=true
Pro
Forms
Slack
Atlassian
Qlik
QuickDB
Azure
Ms Learn
G. Dataset Search
Kaggle
Other favorites
Climbing in Belgium - Survey
Hello Climbers, I ask for some support. Please would you mind filling a 5min anonymous survey about Climbing in Belgium. The purpose is to collect feedbacks for a training project in data analysis (Data analyst - Bruxelles Formation). Deadline to answer is quite short August 31th. If you mind sharing the form to help me getting significant data. Once I defensed my project (Sept 6th), I check whether I can legally publish survey results to whom may be interested. Thank You. Stéphane
Connectez-vous à Google pour enregistrer votre progression. En savoir plus
* Indique une question obligatoire
Date of Birth ? *
Date
mm/dd/yyyy
Where do you live ? (4 digit zipcode only) *
Votre réponse
Demande d'accès en écriture