

دور الاحصاء الرسمي وأهمية التحليل الوصفي في فهم البيانات الضخمة

د . زياد عبدالله



ازدهار البلدان كرامة الإنسان



الأمم المتحدة

الاسكوا
ESCWA

دور الاحصاء الرسمي

الاحصاء الرسمي يلعب دورًا كبيرًا في جمع ونشر البيانات الرسمية

يهدف الاحصاء الرسمي إلى جمع وتوثيق البيانات الاحصائية ذات الصلة بالمجتمع، مثل السكان والاقتصاد والتعليم والصحة وغيرها.

يتم تنفيذه بواسطة الاجهزة الاحصائية الرسمية و الهيئات الاحصائية الوطنية والمنظمات الدولية المعنية بالاحصاء

أهمية التحليل الوصفي في فهم البيانات الضخمة



أهمية التحليل الوصفي في فهم البيانات الضخمة

الأهمية

- التحليل الوصفي هو أداة قوية لفهم البيانات واكتشاف الأنماط والاتجاهات والعلاقات.
- يمكن من اتخاذ قرارات مستنيرة بناءً على البيانات.

التحليل الوصفي للبيانات التقليدية

- يستخدم الإحصاءات البسيطة لتمثيل الاتجاهات الأساسية والتنوع.
- يمكن استخدام الجداول التكرارية والمخططات البيانية لتوفير لمحة سريعة عن البيانات

التحليل الوصفي للبيانات الضخمة

- يتعامل مع كميات هائلة من البيانات المجمعة من مصادر متعددة ومتنوعة.
- يهدف إلى استخراج المعلومات المفيدة من البيانات الأولية.

استعراض أنواع التحليل الوصفي المستخدمة في الاحصاء الرسمي لفهم البيانات الضخمة

التحليل الوصفي الكمي: يشمل استخدام المقاييس الاحصائية الكمية لتفسير وتحليل البيانات الضخمة

من بين أمثلة هذه المقاييس: النزعة المركزية ، مقاييس التشتت، مقاييس الالتواء



مثال: حجم البيانات المستخدمة بالهاتف مقابل الأعمار

العمر (سنوات)	حجم البيانات المستخدمة (جيجابايت)
20	2.5
25	4.2
30	3.8
35	5.1
40	6.2
45	5.5
50	7.0
55	6.8
60	8.5
65	9.3

مثال: حجم البيانات المستخدمة بالهاتف مقابل الأعمار

◦ الوسط (المتوسط 5.89): جيجابايت يمثل متوسط حجم البيانات المستخدمة بالهاتف بين هؤلاء الأفراد.

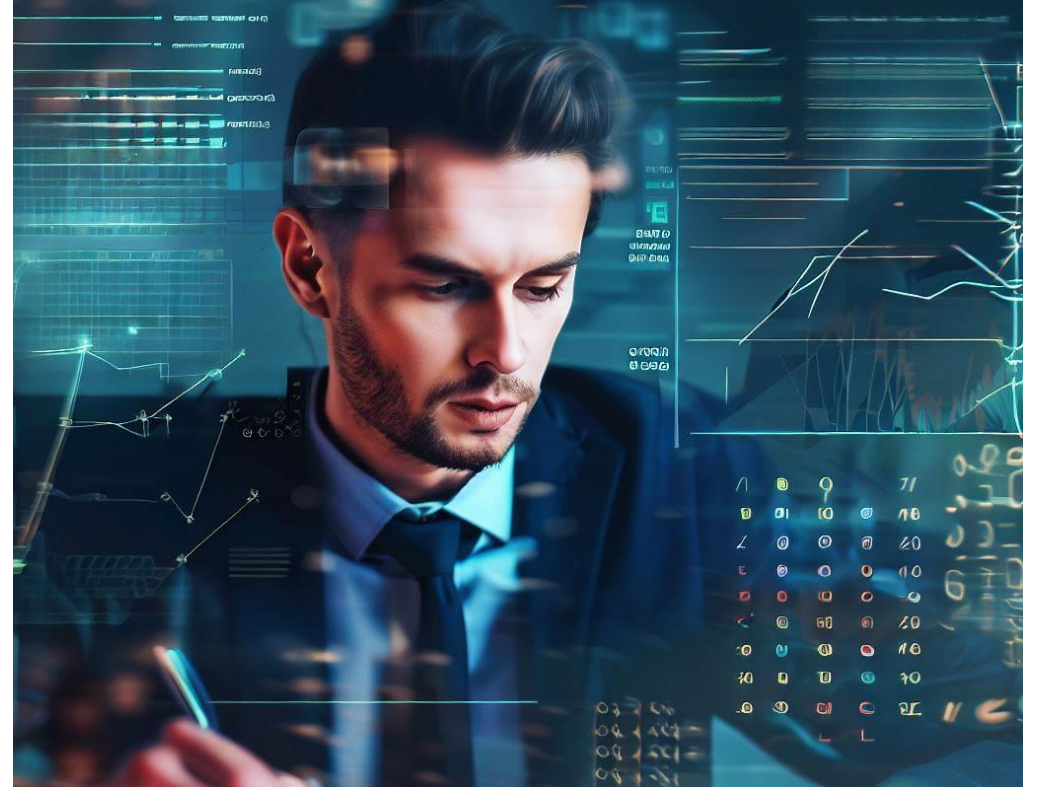
◦ الوسيط: 42.5 سنة يمثل القيمة الموجودة في منتصف الأعمار والتي تفصل البيانات إلى نصفين متساويين.

◦ الانحراف المعياري 2.11: جيجابايت يعطي فكرة عن مدى تباين حجم البيانات بين الأفراد ومدى انحرافها عن المتوسط.

العمر (سنوات)	النتائج	حجم البيانات المستخدمة	النتائج
Mean	42.5	Mean	5.89
Standard Error	4.78	Standard Error	0.667408
Median	42.5	Median	5.85
Mode	#N/A	Mode	#N/A
Standard Deviation	15.13	Standard Deviation	2.11
Kurtosis	-1.2	Kurtosis	-0.56
Skewness	0	Skewness	0.09
Range	45	Range	6.8
Minimum	20	Minimum	2.5
Maximum	65	Maximum	9.3

استعراض أنواع التحليل الوصفي المستخدمة في الاحصاء الرسمي لفهم البيانات الضخمة

- التحليل الوصفي النوعي: يتمحور حول تفسير وفهم البيانات الضخمة من خلال الجوانب النوعية والوصفية.
- يشمل ذلك تحليل المتغيرات النوعية، مثل الجنس، التعليم، وغيرها، وفهم التركيب والتوزيع في البيانات الضخمة.



استعراض أنواع التحليل الوصفي المستخدمة في الإحصاء الرسمي لفهم البيانات الضخمة

◦ استعمل التحليل الوصفي لهذه البيانات

id	time	country	cell_id	event_type	lat	lon
123456	2022-05-10 09:15:00	EG	12345	6	30.0444204	31.2357123
123456	2022-05-10 12:30:00	SA	67890	5	24.7135517	46.6752957
789012	2022-05-10 14:45:00	DZ	98765	1	36.737232	3.0864724
789012	2022-05-11 08:00:00	MA	54321	6	34.0208825	-6.8416503
345678	2022-05-11 10:15:00	JO	24680	1	31.9515694	35.9239625
345678	2022-05-11 14:30:00	EG	13579	5	30.0444204	31.2357123
567890	2022-05-12 16:45:00	SA	98765	6	24.7135517	46.6752957
567890	2022-05-12 18:00:00	DZ	67890	1	36.737232	3.0864724
123456	2022-05-13 09:15:00	EG	54321	6	30.0444204	31.2357123
123456	2022-05-13 12:30:00	MA	13579	5	34.0208825	-6.8416503

مثال على بيانات النظام الآلي لتحديد هوية السفن

FID	mmsi	imo	vessel_name	callsign	vessel_type	vessel_type_code	vessel_type_cargo	vessel_class	length	width	flag_country	flag_code
null	440503000	8815724	55 SHIN YUNG	6MWP	Fishing	30	null	A	55	9	South Korea	440
null	366557000	8419142	MATSON ANCHORAGE	KGTX	Cargo	70	null	A	216	24	USA	366
null	440055000	9019509	ORYONG 325	6MNZ	Fishing	30	null	A	56	10	South Korea	440
null	367542320	null	WALTER L GIBBS	WDG5004	Towing	31	null	A	27	10	USA	367
null	538008215	9844277	OLYMPIC LIFE	V7A2092	Tanker	80	null	A	333	60	Marshall Islands	538
null	345070040	9242106	DONA BLANCA	XCDC	Passenger	60	null	A	22	5	Mexico	345
null	735057514	null	DARWIN	HC2113	Passenger	60	null	A	20	5	Ecuador	735
null	367651380	440	ELK	WDH7758	Cargo	70	null	A	58	15	USA	367
null	366998130	null	TAYLOR MARIE	WDC2822	Tug	52	null	A	22	8	USA	366
null	218791000	9612997	ANTWERPEN EXPRESS	DJCE2	Cargo	79	No Additional Inf...	A	366	48	Germany	218
null	735059299	null	JOLINDA	HC5601	Fishing	30	null	A	45	5	Ecuador	735
null	636016940	9238789	MSC MANU	A8CF3	Cargo	70	null	A	260	32	Liberia	636
null	338392816	null	COOL BREEZE	null	Pleasure Craft	37	null	B	13	5	USA	338
null	636018346	9797187	POLAR CHILE	D5PH8	Cargo	72	Carrying DG,HS or...	A	230	37	Liberia	636
null	563063700	9833541	STI MAGISTER	9V8891	Tanker	80	null	A	183	32	Singapore	563
null	636010032	9018658	SOL DO BRASIL	ELQQ4	Cargo	70	null	A	172	26	Liberia	636
null	338125000	9670339	RUSSELL ADAMS	WDG9047	WIG	20	null	A	81	18	USA	338
null	224559000	8802363	PLAYA DE RODAS	EHQQ	Fishing	30	null	A	55	10	Spain	224
null	316266000	9175298	PLACENTIA PRIDE	VCWB	Tug	52	null	A	38	13	Canada	316
null	710003110	null	PELAGIUS	PR 6983	Tug	52	null	A	30	10	Brazil	710

Vessel ID

Vessel name

Vessels Type

Vessels Size

Vessels Flag

مثال على بيانات النظام الآلي لتحديد هوية السفن

FID	mmsi	imo	vessel_name	callsign	vessel_type	vessel_type_code	vessel_type_cargo	vessel_class	length	width	flag_country	flag_code
null	440503000	8815724	55 SHIN YUNG	6MNP	Fishing	30	null	A	55	9	South Korea	440
null	366557000	8419142	MATSON ANCHORAGE	KGTX	Cargo	70	null	A	216	24	USA	366
null	440055000	9019509	ORYONG 325	6MNZ	Fishing	30	null	A	56	10	South Korea	440
null	367542320	null	WALTER L GIBBS	WDG5004	Towing	31	null	A	27	10	USA	367
null	538008215	9844277	OLYMPIC LIFE	V7A2092	Tanker	80	null	A	333	60	Marshall Islands	538
null	345070040	9242106	DONA BLANCA	XCDC	Passenger	60	null	A	22	5	Mexico	345
null	735057514	null	DARWIN	HC2113	Passenger	60	null	A	20	5	Ecuador	735
null	367651380	440	ELK	WDH7758	Cargo	70	null	A	58	15	USA	367
null	366998130	null	TAYLOR MARIE	WDC2822	Tug	52	null	A	22	8	USA	366
null	218791000	9612997	ANTWERPEN EXPRESS	DJCE2	Cargo	79	No Additional Inf...	A	366	48	Germany	218
null	735059299	null	JOLINDA	HC5601	Fishing	30	null	A	45	5	Ecuador	735
null	636016940	9238789	MSC MANU	A8CF3	Cargo	70	null	A	260	32	Liberia	636
null	338392816	null	COOL BREEZE	null	Pleasure Craft	37	null	B	13	5	USA	338
null	636018346	9797187	POLAR CHILE	DSPH8	Cargo	72	Carrying DG,HS or...	A	230	37	Liberia	636
null	563063700	9833541	STI MAGISTER	9V8891	Tanker	80	null	A	183	32	Singapore	563
null	636010032	9018658	SOL DO BRASIL	ELQQ4	Cargo	70	null	A	172	26	Liberia	636
null	338125000	9670339	RUSSELL ADAMS	WDG9047	WIG	20	null	A	81	18	USA	338
null	224559000	8802363	PLAYA DE RODAS	EHQQ	Fishing	30	null	A	55	10	Spain	224
null	316266000	9175298	PLACENTIA PRIDE	VCWB	Tug	52	null	A	38	13	Canada	316
null	710003110	null	PELAGIUS	PR 6983	Tug	52	null	A	30	10	Brazil	710

Vessel ID

Vessel name

Vessels Type

Vessels Size

Vessels Flag

مثال على بيانات النظام الآلي لتحديد هوية السفن

destination	eta	draught	position	longitude	latitude	sog	cog	rot	heading	nav_status	nav_status_code
null	null	0.0	POINT (13.1726333... -164.43488333	-164.43488333	13.17263333	3.7	116.8	0.0	0	Under Way Using E...	0
TACOMA WA	null	9.0	POINT (53.9401883... -164.57464667	-164.57464667	53.94018833	19.3	86.8	16.11514409	86	Under Way Using E...	0
null	null	3.7	POINT (1.6708 -15... -153.56116667	-153.56116667	1.6708	4.0	152.6	0.0	0	Under Way Using E...	0
HOUSTON	null	2.9	POINT (29.7433333... -94.08	-94.08	29.74333333	5.0	230.0	0.0	0	Unknown	16
GALVESTON	null	11.0	POINT (28.3352133... -93.05576667	-93.05576667	28.33521333	11.5	303.8	0.0	302	Under Way Using E...	0
null	null	0.0	POINT (18.6533333... -91.84166667	-91.84166667	18.65333333	0.0	212.0	0.0	0	Not Defined	15
CRUCEROS INTERISLAS	null	0.0	POINT (-0.75 -90.31)	-90.31	-0.75	0.0	276.0	0.0	0	At Anchor	1
FOURCHON	null	4.0	POINT (28.35 -90... -90.66666667	-90.66666667	28.35	0.0	26.0	0.0	0	Under Way Using E...	0
US^OEWS>OE70	null	2.8	POINT (30.0466666... -90.6	-90.6	30.04666667	0.0	173.0	0.0	0	Unknown	16
KRPUS	null	12.9	POINT (8.24166666... -86.84666667	-86.84666667	8.24166667	19.0	284.0	0.0	0	Under Way Using E...	0
FAENA D PESCA	null	0.0	POINT (-11.474056... -84.07834	-84.07834	-11.47405667	0.0	0.0	0.0	129	Engaged In Fishing	7
PAROD	null	8.8	POINT (-0.11949 -... -81.113605	-81.113605	-0.11949	17.9	13.5	0.0	13	Under Way Using E...	0
null	null	0.0	POINT (26.16917 -... -80.10563	-80.10563	26.16917	0.0	0.0	0.0	0	Unknown	16
BALBOA	null	10.2	POINT (-33.592733... -71.61748333	-71.61748333	-33.59273333	0.0	222.2	0.0	181	Moored	5
BR SLZ	null	12.2	POINT (14.603895 ... -68.09905	-68.09905	14.603895	11.3	114.2	0.0	116	Under Way Using E...	0
US ILG	null	9.4	POINT (26.2783333... -64.30666667	-64.30666667	26.27833333	17.0	312.0	0.0	0	Under Way Using E...	0
GT GUY	null	4.2	POINT (6.78647833... -58.17381333	-58.17381333	6.78647833	0.0	46.0	0.0	13	Moored	5
FISHING GROUND	null	7.2	POINT (-35.757288... -55.027085	-55.027085	-35.75728833	10.6	131.1	0.0	131	Moored	5
null	null	0.0	POINT (47.7732133... -54.01134167	-54.01134167	47.77321333	0.0	49.0	0.0	8	Not Defined	15
SAO LUIS	null	4.0	POINT (-2.59382 -... -44.36726833	-44.36726833	-2.59382	0.1	208.6	0.0	0	Underway Sailing	8

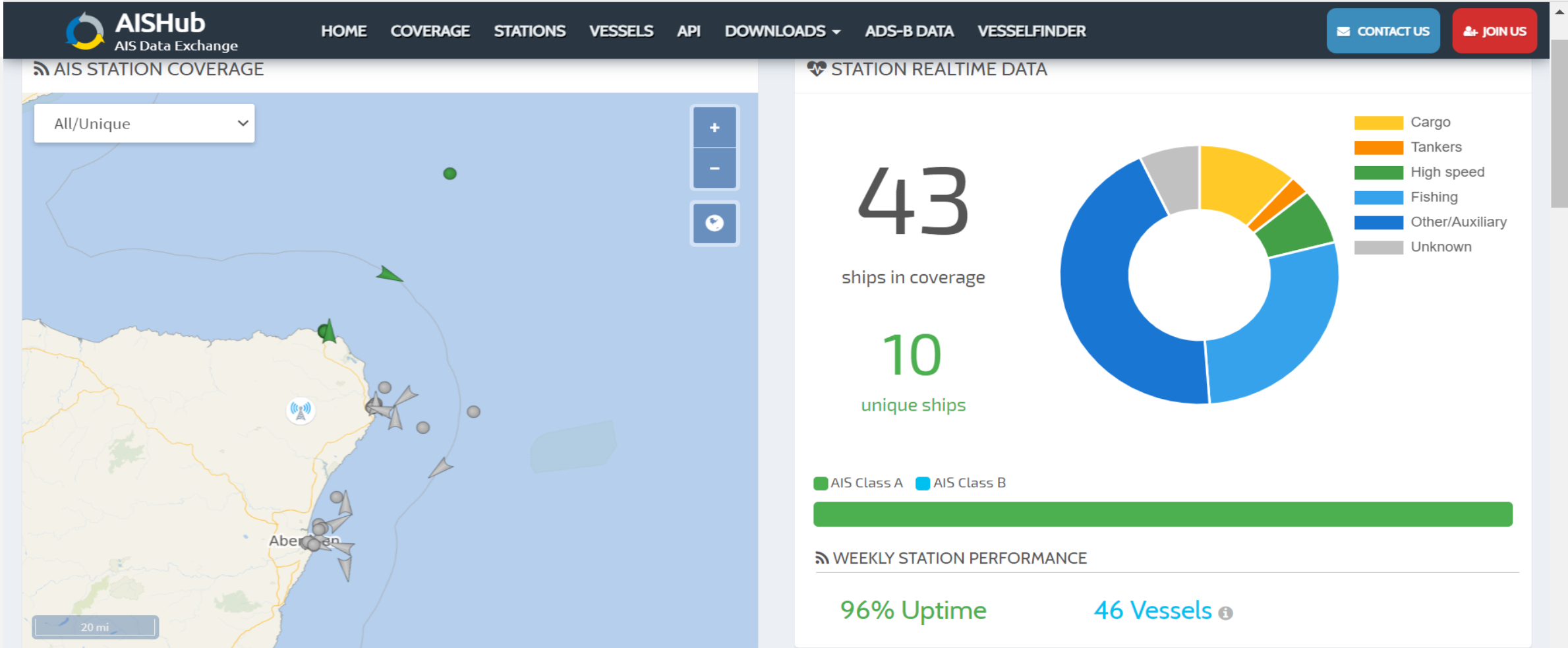
Destination

Geospatial
location

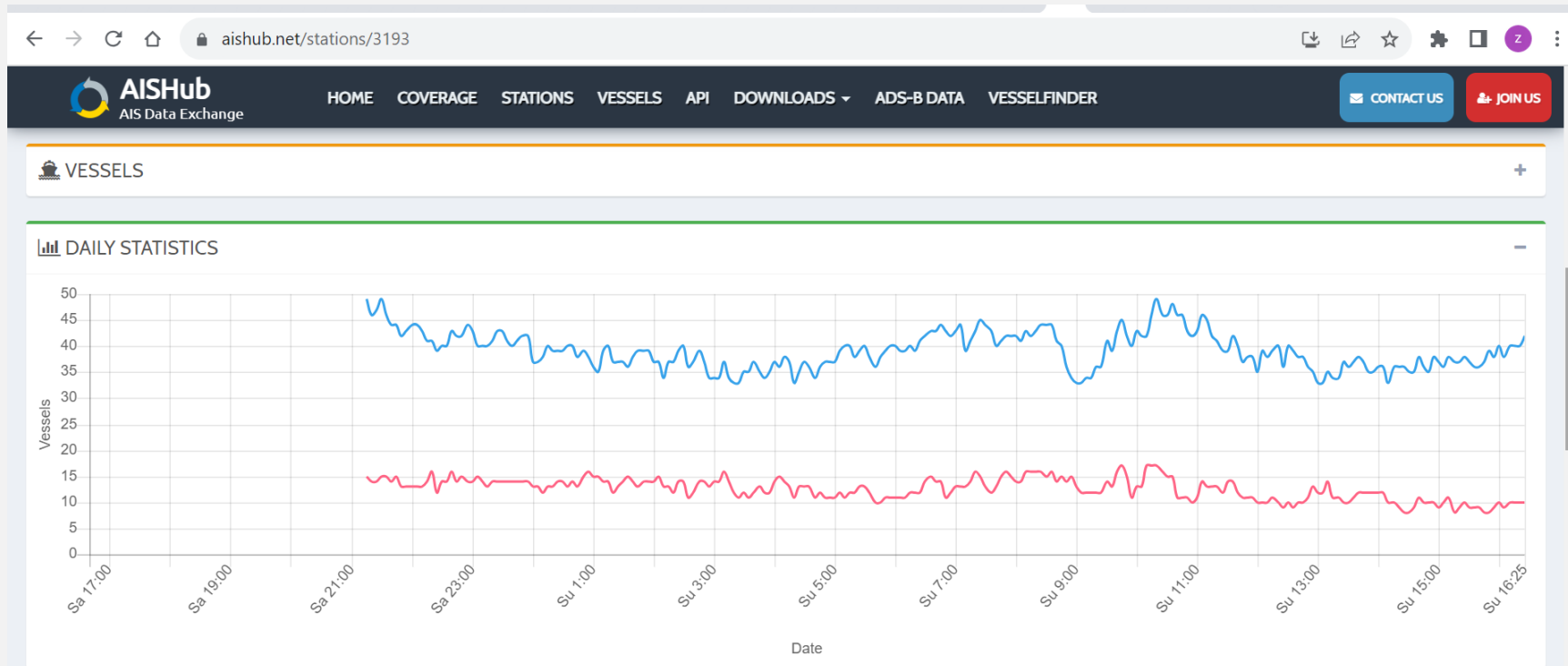
Speed

Navigation
Status

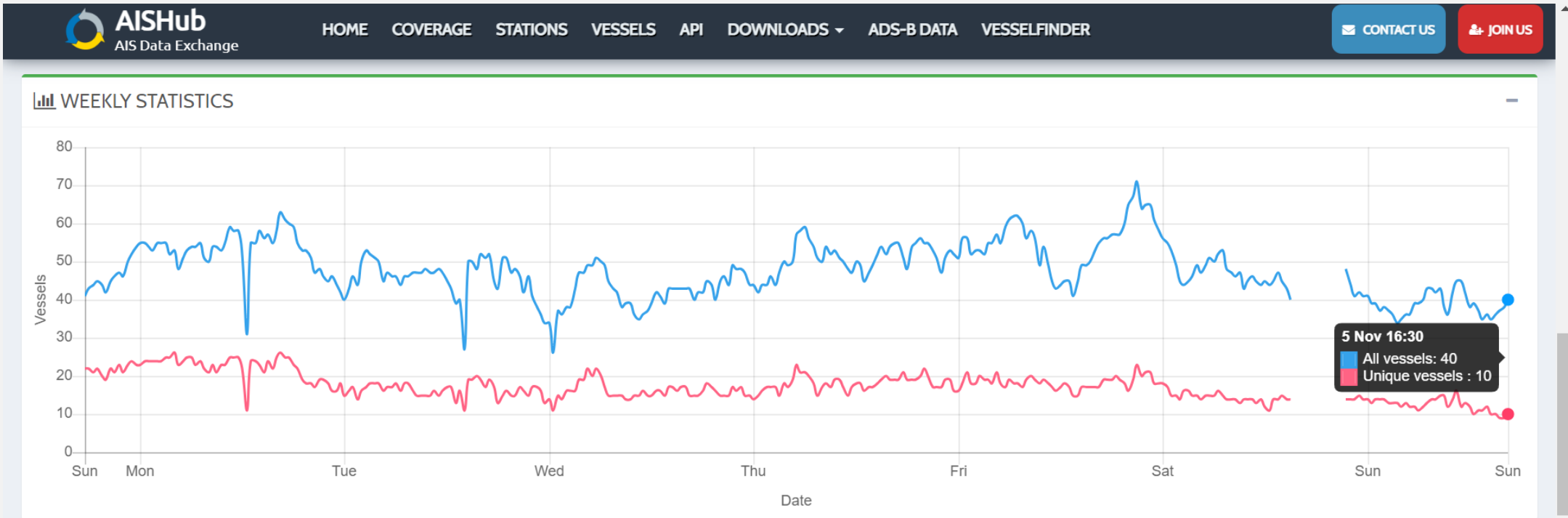
مثال على بيانات النظام الآلي لتحديد هوية السفن



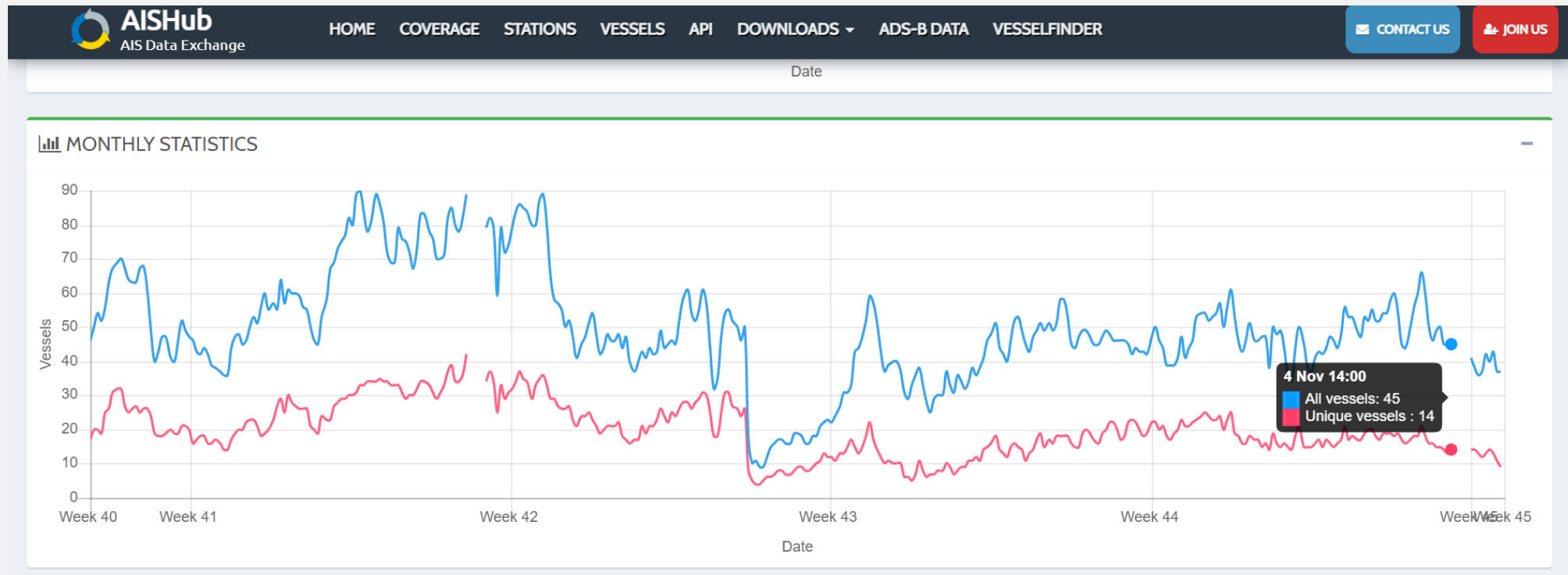
مثال على بيانات النظام الآلي لتحديد هوية السفن



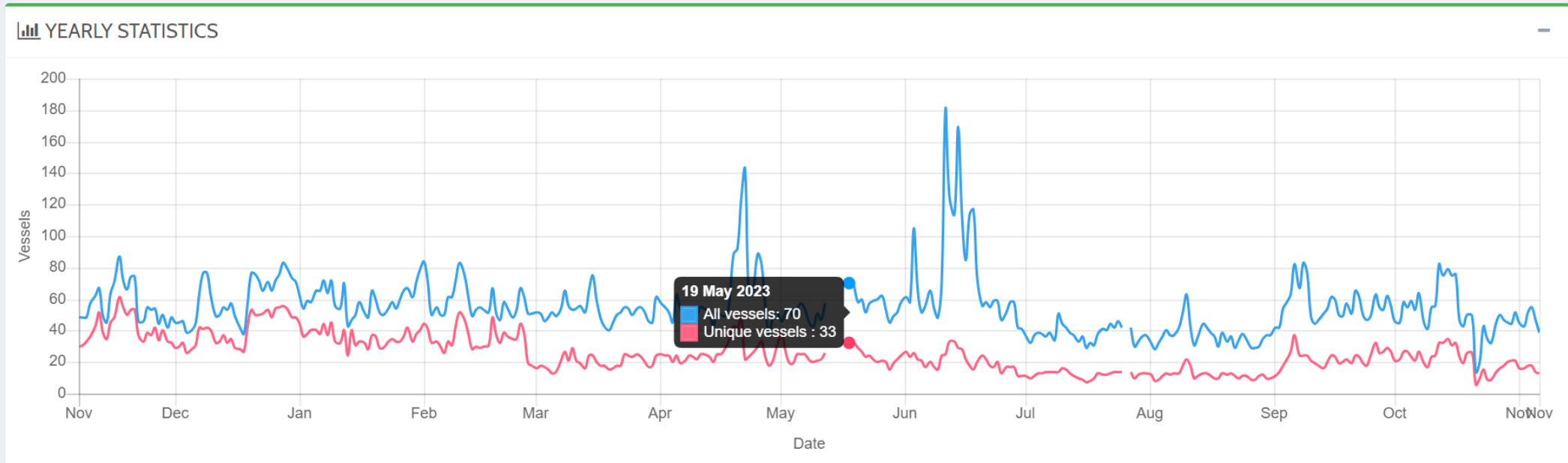
مثال على بيانات النظام الآلي لتحديد هوية السفن



مثال على بيانات النظام الآلي لتحديد هوية السفن



مثال على بيانات النظام الآلي لتحديد هوية السفن



مثال: استعمل التحليل الزمني للبيانات التالية

id	call_time
123456	2023-08-27 10:15:00
234567	2023-08-27 13:30:00
123456	2023-08-28 09:45:00
345678	2023-08-28 12:00:00
234567	2023-08-29 15:20:00
...	...

مثال: تحليل الزمني لنشاط المكالمات الهاتفية خلال أيام الأسبوع

تحويل التواريخ إلى أيام الأسبوع:
تحويل التواريخ المسجلة إلى أيام الأسبوع (مثل الأحد، الاثنين، ...، السبت).
تجميع البيانات حسب أيام الأسبوع:
حساب عدد المكالمات في كل يوم من أيام الأسبوع.
إنشاء رسم بياني للأنماط الزمنية
استخدم رسم بياني مثل الرسم الشريطي أو الرسم الخطي لتوضيح عدد المكالمات خلال أيام الأسبوع.

التحليل المتعدد المتغيرات

- **Multivariate Analysis:** It's a statistical method that examines multiple variables to understand the relationships between them.
- **Factor Analysis:** This technique is used to uncover the latent structure (dimensions) of a set of variables. It reduces the attribute space from a larger number of variables to a smaller number of factors.
- **Multiple Regression Analysis:** This technique is used to study the relationship between one variable (the dependent variable) and another (the independent variable), taking into account the effect of other variables.
- **Big Data Analysis:** Multivariate analysis techniques are used in big data analysis to discover patterns, trends, and relationships in the data.

• **التحليل المتعدد المتغيرات:** هو طريقة إحصائية تدرس متغيرات متعددة لفهم العلاقات بينها.

• **تحليل العوامل:** يتم استخدام هذه التقنية للكشف عن الهيكل الكامن (الأبعاد) لمجموعة من المتغيرات. يقلل من مساحة السمة من عدد أكبر من المتغيرات إلى عدد أصغر من العوامل.

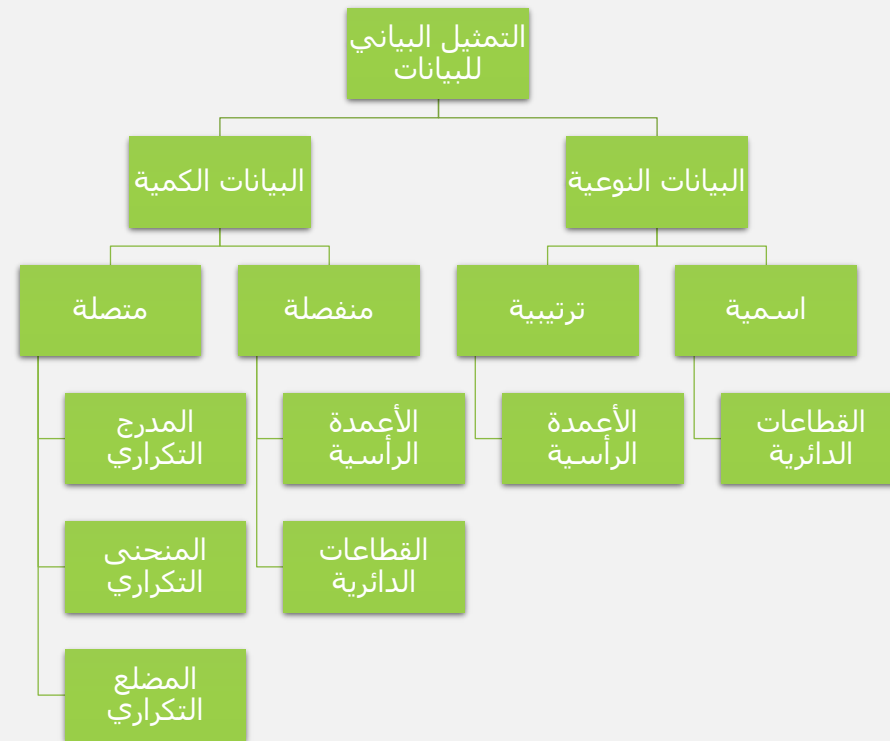
• **تحليل الانحدار المتعدد:** يتم استخدام هذه التقنية لدراسة العلاقة بين متغير (المتغير المستقل) ومتغير آخر (المتغير التابع)، مع الأخذ في الاعتبار تأثير المتغيرات الأخرى.

• **تحليل البيانات الكبيرة:** يتم استخدام تقنيات التحليل المتعدد المتغيرات في تحليل البيانات الكبيرة لاكتشاف الأنماط والاتجاهات والعلاقات في البيانات.

استعمل التحليل المتعدد المتغيرات في البيانات التالية:

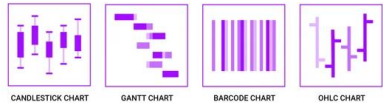
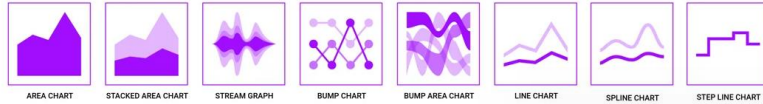
مستخدم	الجنس	العمر	نوع الحدث	حجم البيانات المستخدمة
1	ذكر	30	6	3.5
2	أنثى	25	5	2.7
3	ذكر	40	1	4.2
4	أنثى	28	6	5.1
5	ذكر	32	1	3.8
6	ذكر	22	5	2.5
7	أنثى	29	6	3.0
8	ذكر	35	1	4.8
9	ذكر	31	6	3.9
10	أنثى	27	1	4.5

عرض البيانات

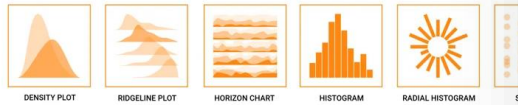


عرض البيانات

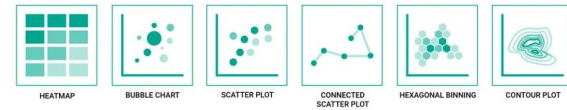
DATA OVER TIME (TEMPORAL)



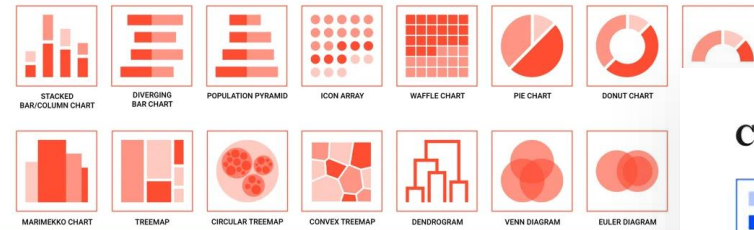
DISTRIBUTION



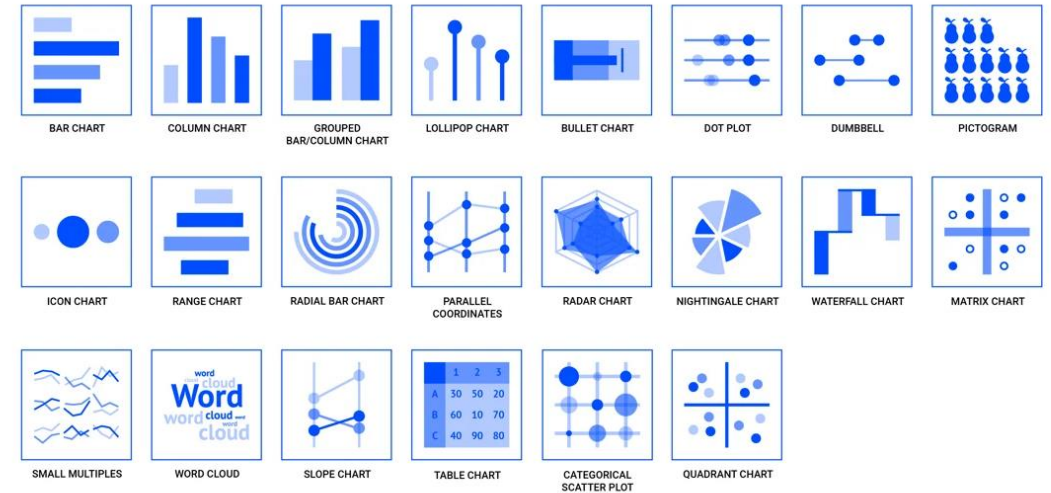
CORRELATION



PART-TO-WHOLE & HIERARCHICAL



COMPARISON



التحديات لتحليل الوصفي في البيانات الضخمة

حجم البيانات: تعد البيانات الضخمة ذات حجم هائل وقد تصعب معالجتها وتحليلها بالطرق التقليدية. يتطلب التحليل الوصفي في البيانات الضخمة استخدام تقنيات وأدوات تحليل متقدمة للتعامل مع حجم البيانات الكبير.

التعقيد والتنوع: يمكن أن تكون البيانات الضخمة معقدة ومتنوعة، مما يزيد التحدي في استخلاص الأنماط والتفسيرات. قد تحتاج إلى تطبيق تقنيات تحليل متعددة لاستخراج المعلومات القيمة والمفهومة من البيانات المتنوعة.

جودة البيانات: يجب أن يتم التعامل مع جودة البيانات في البيانات الضخمة، بما في ذلك التحقق من البيانات المفقودة والقيم الشاذة في البيانات. قد تحتاج إلى تطبيق إجراءات تنقية البيانات والمعالجة لضمان جودة البيانات.

المزايا لتحليل الوصفي في البيانات الضخمة

الاكتشافات والأنماط: يمكن لتحليل الوصفي في البيانات الضخمة أن يساهم في اكتشاف الأنماط والتقاربات في البيانات. يمكن استخدام تقنيات التجميع والتصنيف والرسوم البيانية لتحليل البيانات وتوضيح الأنماط والعلاقات الموجودة.

دقة وموثوقية النتائج: باستخدام التحليل الوصفي في البيانات الضخمة، يمكن الحصول على نتائج دقيقة وموثوقة بناءً على حجم العينة الكبير وتمثيلية البيانات. يمكن أن يؤدي ذلك إلى اتخاذ قرارات أفضل وتوجيه السياسات العامة بشكل أكثر دقة.

التحليل الزمني والمكاني: يمكن استخدام التحليل الوصفي في البيانات الضخمة لتحليل العلاقات الزمنية والمكانية في البيانات. يمكن استخدام تقنيات التحليل الزمني والتحليل المكاني لاستكشاف الأنماط والتغيرات عبر الزمن وتحليل العلاقات المكانية بين المتغيرات.

استنتاج المحاضرة

التحليل الوصفي في البيانات الضخمة يلعب دورًا حيويًا في تطوير الاحصاء الرسمي

فهم البيانات الضخمة: يساعد التحليل الوصفي في فهم البيانات الضخمة وتوصيفها بشكل مفصل وشامل. يتيح استخدام أدوات وتقنيات التحليل الوصفي فهم الأنماط والعلاقات بين المتغيرات والتحليل المفصل للتغيرات عبر الزمن والمكان.

استخلاص المعلومات القيمة: يمكن استخدام التحليل الوصفي في البيانات الضخمة لاستخراج المعلومات القيمة والمفهومة. يتيح تحليل البيانات الكمية والنوعية فهم التوزيعات والأنماط والاختلافات في البيانات، مما يساهم في اتخاذ القرارات الأساسية.

استنتاج المحاضرة

دعم اتخاذ القرارات: يمكن أن يسهم التحليل الوصفي في البيانات الضخمة في اتخاذ القرارات الرشيدة. من خلال فهم العلاقات والأنماط والتغيرات في البيانات، يتيح التحليل الوصفي لصانعي القرار تحليل الأداء وتحديد المجالات التي تحتاج إلى تحسين وتوجيه السياسات والاستراتيجيات بناءً على الأدلة القائمة على البيانات.

تحسين الاحصاء الرسمي: يمكن استخدام التحليل الوصفي في البيانات الضخمة لتحسين الاحصاء الرسمي بشكل عام. يتيح التحليل الوصفي فهم العلاقات الجماعية والتغيرات الزمنية والمكانية، وبالتالي يسهم في تحديث العمليات والمنهجيات الإحصائية وتعزيز جودة البيانات وتحسين دقة الإحصاء الرسمي.



ازدهار البلدان كرامة الإنسان



شكرا لكم