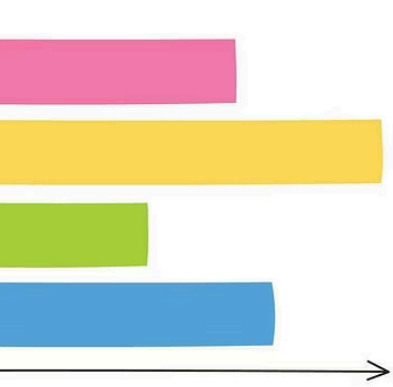




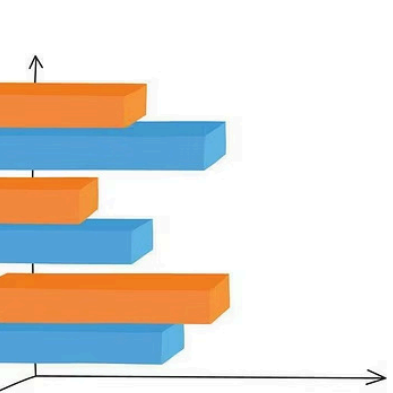
PCCI DATA – READMIT30 PROFILE

Summary of CSV: 28 variables, 12,980 observations, 0 missing cells. Total memory
~1.64 MB. Average record size 0.13 kB.

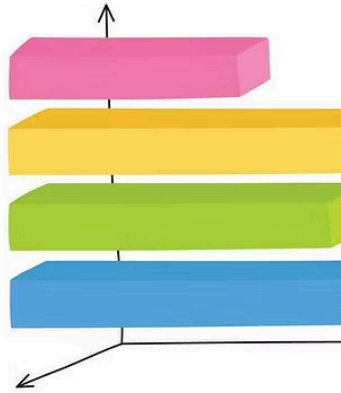
BAR CHART



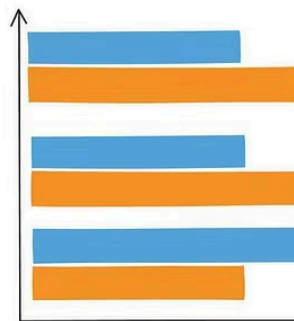
Stacked Bar



3-D Clustered Bar



3-D Stacked Bar



Clustered Bar

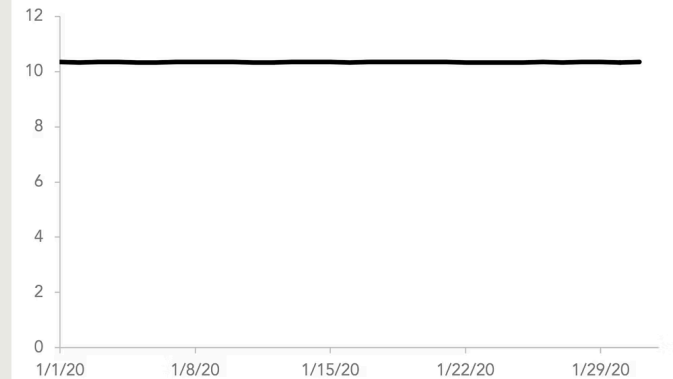
VARIABLE TYPES (28)

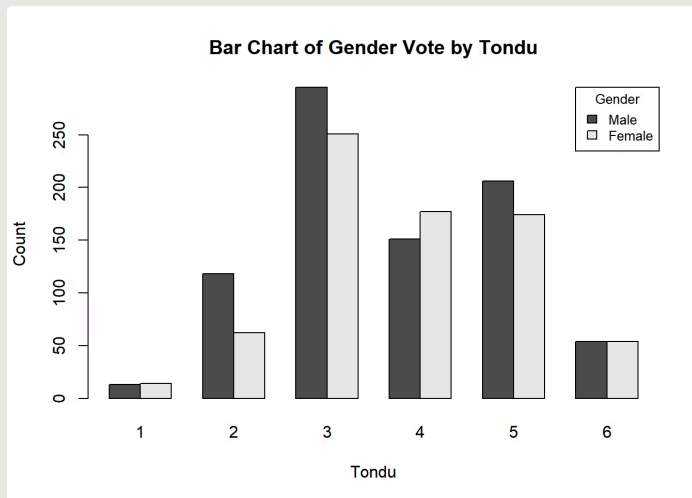
All 28 variables are numeric in the exported profile: Patient_Id, Gender, Mental_Status, Insurance_Provider, Tobacco_User, Depression, ICU, Drug_Abuse, Mood_Disorder, Diabetes, Anxiety, Obesity, Dementia, Age, Bmi, Weight, Height, Pulse, Temperature, Pat_Pain_Score, ER_Visits, IP_Visits, Chronic_Conditions, Glucose, Condition, Care_Plan_Following_Discharge, Cost_Of_Initial_Stay, Readmit30.

OVERALL VALUE DISTRIBUTION

Distribution shows a roughly flat frequency (~600) across values with a sharp drop near the maximum (15821). Statistics excerpt: Avg 7904.54, Std 4555.84, Min 1, Max 15821.

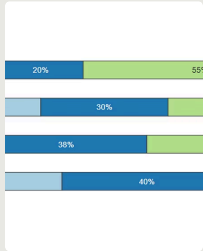
ABC Corporation January daily valuation
IN BILLIONS (USD)





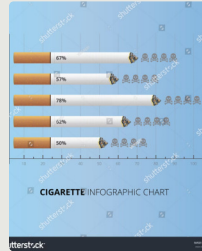
GENDER & MARITAL STATUS

Gender: F \approx 6,000, M \approx 6,500. Marital_Status: Single \approx 6,500, Married \approx 6,000.



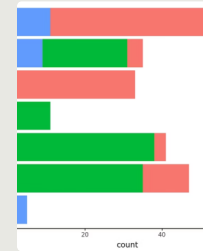
INSURANCE PROVIDER

Medicare ≈ 10,000;
 Commercial ≈ 1,500; Medicaid ≈ 500.



TOBACCO USER

Quit ≈ 7,000; Never ≈ 4,000;
 Yes ≈ 2,000; Not Asked ≈ 500;
 Passive ≈ 100.



MENTAL HEALTH FLAGS

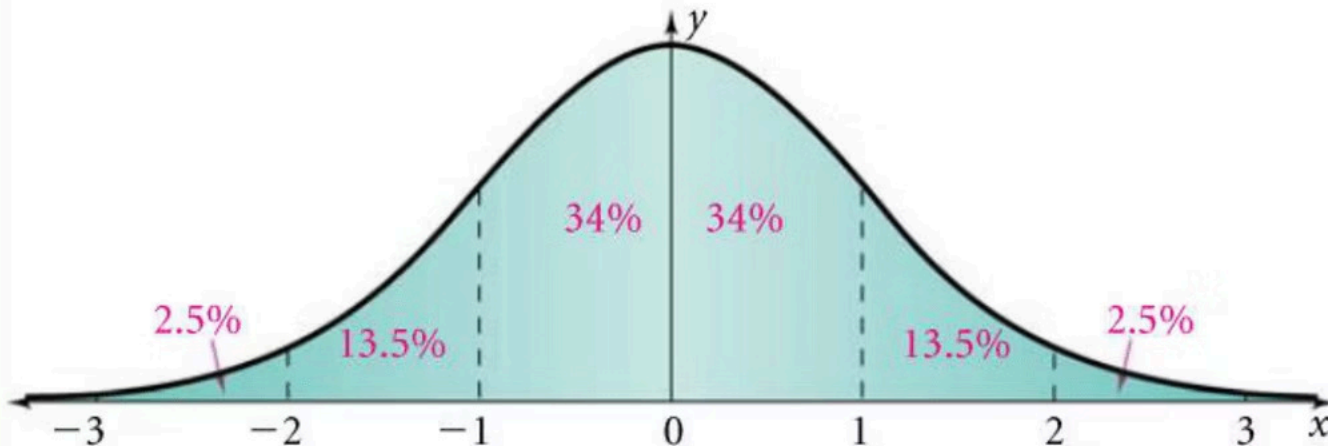
Depression: No ≈ 12,000, Yes ≈ 1,000. Similar patterns observed for ICU, Drug_Abuse, Mood_Disorder, Diabetes (mostly No).

The Bell Curve

Normal Distribution – shows data that vary randomly from the mean.

Normal Curve – The pattern the data form is a bell-shaped curve.

The Standard Normal Bell Curve



68% of the data fall within one standard deviation of the mean

95% of the data fall within two standard deviations of the mean

KEY NUMERIC SUMMARIES

Age: Avg 71.62 (Min 18.1, Max 90). Bmi: Avg 30.34 (Min 14.0, Max 54.47). Weight: Avg 190.99 (Min 42, Max 392). Pulse: Avg 78.68. Temperature: Avg 97.72.

CHRONIC CONDITIONS

Avg 6.61; distribution peaks at 1 (~2,500) and 2 (~2,200); values range 1–27.

GLUCOSE

Avg 144.63, right-skewed distribution, range 58–423, many unique values (356).

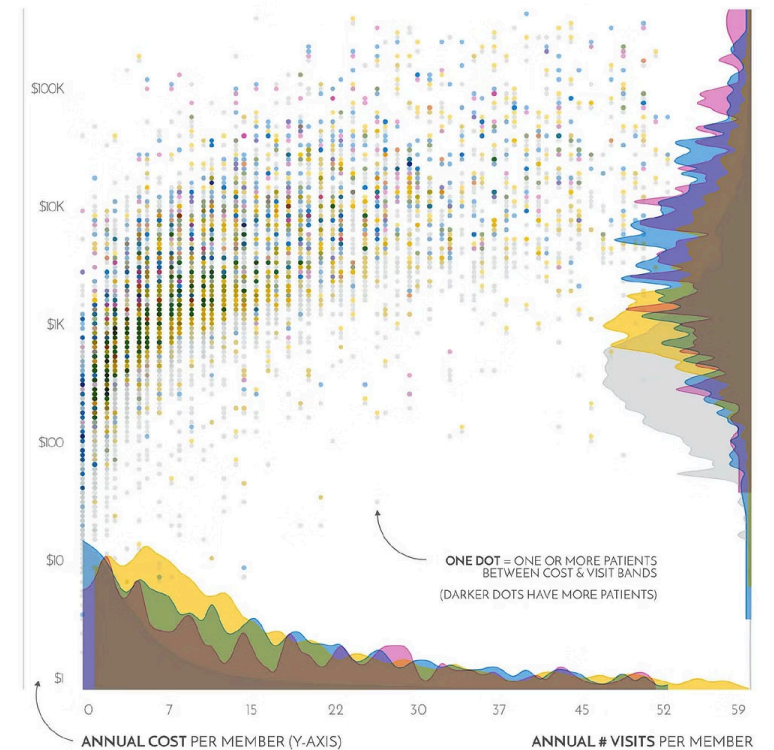
COST OF INITIAL STAY

Avg \$7,268.54, right-skewed, Min \$9.40, Max \$117,843.12, Std ≈ 6,171.89.

Utilization density plot

Cost and utilization of 50K ACO patients

VISITS PER MEMBER	PORTION OF POPULATION	ANNUAL COST PER MEMBER
7.94	BASELINE	\$4,488
16.31	DEPRESSION	\$17,205
15.41	HEAVY ED USERS	\$16,742
16.46	DRUG & ALCOHOL ABUSE	\$24,799



ARCADIA™

© 2017 arcadia.io

Made with GAMMA



DATA SNAPSHOT

KEY DISTRIBUTIONS

TARGET FLAG

READMIT30 (TARGET)

Binary: 0 \approx 10,000, 1 \approx 1,500. Avg 0.19. Use for readmission modeling and outcome analysis.

CheXpert [15]	X-Ray	Lung	~900x1300	234,316624	2	CLS
CheXRay 2017 [19]	X-Ray	Lung	~1000x700	5,232624	2	CLS
LUNA [25]	CT	Lung	512x512	207	2	SEG
CHAOS-CT [18]	CT	Liver	512x512	2874	2	SEG
NH1 CT-SEG [21]	CT	Pancreas	512x512	7,141	2	SEG
CHAOS-MRI [18]	MRI	Liver/Kidney/Spleen	320x320	992	5	SEG
CardiacMRI [2]	MRI	Heart	226x226	399	3	SEG
ACDC [1]	MRI	Heart	320x320	-	4	SEG
TCTA [14, 5]	MRI	Prostate	320x320	-	3	SEG
PROMISE12 [23]	MRI	Prostate	-	-	-	SEG
ISC2019 [4]	Dermoscopy	Skin	1024x768	25,331	9	CLS
TNBC [27]	HE Stained	Cell	512x512	50	2	SEG
GaS [9]	HE Stained	Cell	775x522	8580	2	SEG
MaNa [22]	HE Stained	Cell	1000x1000	2014	2	SEG
BC-2015 [1]	HE Stained	Breast cancer	2000x1536	269	2	CLS
DRISHFI-GS [11]	Fundoscopic	Optic Disc, Cup	2045x1752	5051	2	SEG
DRIONS-OR [7]	Fundoscopic	Fundus	400x400	9960	2	SEG
RIM-e3 [13]	Fundoscopic	Fundus	1072x1072	9960	2	SEGCLS
REFUGE [26]	Fundoscopic	Optic Disc, Cup/Glaucoma	1634x1634	400400	2/2	SEGCLS
IRF [8]	Fundoscopic	Vessel/Glaucoma	3584x3336	45	3	SEGCLS
DRIVE [37]	Fundoscopic	Vessel	565x564	2020	2	SEG
MESSIDOR [11]	Fundoscopic	Diabetic Retinopathy	2240x1488	1,200	4/4	CLS
STARE [14]	Fundoscopic	Diabetic Retinopathy/Vessel	700x650	397	15	SEGCLS
EyesPC5 [16]	Fundoscopic	Diabetic Retinopathy	4000x4000	35126	5	CLS
BUS [38]	Ultrasound	Breast lesion	500x500	163	2	SEG
BUS1 [1]	Ultrasound	Breast lesion	500x500	780	3	SEGCLS
OCT2017 [17]	OCT	Fundus	~512x512	108,3091,000	4	CLS
KIMICS [37]	Crossscope	Cervical	~3000x3000	1,463	3	CLS
CIFAR100 [20]	Natural Image	-	32x32	50,000/10,000	100	CLS
ImageNet [21]	Natural Image	-	~500x375	1.3M/60000	1000	CLS

FIRST 5 RECORDS (PREVIEW)

Shows typical rows: Patient_Id, demographics, flags (Depression, ICU), Visits, Care plan, Cost, Readmit30.



PROFILING TOOL

StatSim Profile is free, client-side, open-source; processes large CSVs locally and produces summaries, charts, and top values.



NEXT STEPS & RESOURCES

- Use profile to pick features (age, chronic count, glucose, cost, prior visits).
- Address skewed costs and glucose with transforms or robust models.
- Consider stratified sampling for imbalanced Readmit30 ($\approx 19\%$ positive).
- Explore StatSim Preview / Vis for PCA, t-SNE, UMAP.

StatSim Profile – free, client-side tool for quick CSV profiling and visualization.