

Sinha, Aakriti

**CS - 1 Quantitative Analysis and Modeling for
S&P 500 [Noida]: Metadata Description (2013-2018)**

Study Documentation

October 8, 2020

Metadata Production

Identification	CS-1-Asin
----------------	-----------

Table of Contents

Overview.....	4
Scope & Coverage.....	4
Producers & Sponsors.....	4
Data Collection.....	4
Files Description.....	5
cs-1.....	5
Variables List.....	6
cs-1.....	6
Variables Description.....	7
cs-1.....	8

CS - 1 Quantitative Analysis and Modeling for S&P 500 [Noida]: Metadata Description (2013-2018)

Overview	
Identification	CS1-Asin-D
Version	1.0-081020
<p><u>Abstract</u></p> <p>Stock market data provides interesting opportunities for analysis on trends and forecasts. This dataset pertains to historical stock data of all S&P 500 companies for 2013-2018. It has daily data for individual stocks labelled by ticker name. The data is useful in quantitative trading analysis to measure stock performance. The current case study utilizes this dataset for four main objectives: 1) establish a weekly volatility index, 2) explore key statistical tendencies of a specific stock, 3) pairs trading, and 4) prediction of whether closing price will be higher or lower than opening price. The major variables present are:</p> <p>Date - in format: yy-mm-dd</p> <p>Open - price of the stock at market open (in USD)</p> <p>High - highest price reached in the day (in USD)</p> <p>Low - lowest price reached in the day (in USD)</p> <p>Close - price of the stock at market close (in USD)</p> <p>Volume - number of shares traded</p> <p>Name - ticker name of the stock</p>	
Unit of Analysis	Company Stock

Scope & Coverage	
Keywords	S&P 500, Stock Market
<u>Countries</u>	
<p><u>Geographic Coverage</u></p> <p>USA</p>	
<p><u>Universe</u></p> <p>Companies listed in S&P 500 Index</p>	

Producers & Sponsors	
Primary Investigator(s)	Sinha, Aakriti
Other Producer(s)	CogniTensor Technology Private Limited , Collection Cam Nugent , www.kaggle.com , Original Producer

Data Collection	
Data Collection Mode	Data acquired from The Investor's Exchange api using a script found at https://github.com/CNuge/kaggle_code/blob/master/stock_data/getSandP.py

Files Description

Dataset contains 1 file(s)

cs-1	
# Cases	619040
# Variable(s)	7

Variables List

Dataset contains 7 variable(s)

File cs-1							
#	Name	Label	Type	Format	Valid	Invalid	Question
1	date	date	discrete	character-10	619040	0	-
2	open	open	continuous	numeric.2	619029	11	-
3	high	high	continuous	numeric.2	619032	8	-
4	low	low	continuous	numeric.2	619032	8	-
5	close	close	continuous	numeric.2	619040	0	-
6	volume	volume	continuous	numeric.0	619040	0	-
7	Name	Name	discrete	character-5	619040	0	-

Variables Description

Dataset contains 7 variable(s)

File : cs-1

# date: date				
Information		[Type= discrete] [Format=character] [Missing=*]		
Statistics [NW/ W]		[Valid=619040 /-] [Invalid=0 /-]		
# open: open				
Information		[Type= continuous] [Format=numeric] [Range= 1.62-2044] [Missing=*]		
Statistics [NW/ W]		[Valid=619029 /-] [Invalid=11 /-] [Mean=83.023 /-] [StdDev=97.379 /-]		
# high: high				
Information		[Type= continuous] [Format=numeric] [Range= 1.69-2067.99] [Missing=*]		
Statistics [NW/ W]		[Valid=619032 /-] [Invalid=8 /-] [Mean=83.778 /-] [StdDev=98.208 /-]		
# low: low				
Information		[Type= continuous] [Format=numeric] [Range= 1.5-2035.11] [Missing=*]		
Statistics [NW/ W]		[Valid=619032 /-] [Invalid=8 /-] [Mean=82.256 /-] [StdDev=96.507 /-]		
# close: close				
Information		[Type= continuous] [Format=numeric] [Range= 1.59-2049] [Missing=*]		
Statistics [NW/ W]		[Valid=619040 /-] [Invalid=0 /-] [Mean=83.044 /-] [StdDev=97.39 /-]		
# volume: volume				
Information		[Type= continuous] [Format=numeric] [Range= 0-618237630] [Missing=*]		
Statistics [NW/ W]		[Valid=619040 /-] [Invalid=0 /-] [Mean=4321823.396 /-] [StdDev=8693609.512 /-]		
# Name: Name				
Information		[Type= discrete] [Format=character] [Missing=*]		
Statistics [NW/ W]		[Valid=619040 /-] [Invalid=0 /-]		
Value	Label	Cases	Percentage	
A		1259		0.2%
AAL		1259		0.2%
AAP		1259		0.2%
AAPL		1259		0.2%
ABBV		1259		0.2%
ABC		1259		0.2%
ABT		1259		0.2%
ACN		1259		0.2%
ADBE		1259		0.2%
ADI		1259		0.2%
ADM		1259		0.2%
ADP		1259		0.2%
ADS		1259		0.2%
ADSK		1259		0.2%
AEE		1259		0.2%
AEP		1259		0.2%
AES		1259		0.2%
AET		1259		0.2%
AFL		1259		0.2%

File : cs-1

Name: Name

Value	Label	Cases	Percentage
AGN		1259	0.2%
AIG		1259	0.2%
AIV		1259	0.2%
AIZ		1259	0.2%
AJG		1259	0.2%
AKAM		1259	0.2%
ALB		1259	0.2%
ALGN		1259	0.2%
ALK		1259	0.2%
ALL		1259	0.2%
ALLE		1063	0.2%
ALXN		1259	0.2%
AMAT		1259	0.2%
AMD		1259	0.2%
AME		1259	0.2%
AMG		1259	0.2%
AMGN		1259	0.2%
AMP		1259	0.2%
AMT		1259	0.2%
AMZN		1259	0.2%
ANDV		1259	0.2%
ANSS		1259	0.2%
ANTM		1259	0.2%
AON		1259	0.2%
AOS		1259	0.2%
APA		1259	0.2%
APC		1259	0.2%
APD		1259	0.2%
APH		1259	0.2%
APTV		44	0.0%
ARE		1259	0.2%
ARNC		1259	0.2%
ATVI		1259	0.2%
AVB		1259	0.2%
AVGO		1259	0.2%
AVY		1259	0.2%
AWK		1259	0.2%
AXP		1259	0.2%
AYI		1259	0.2%
AZO		1259	0.2%
BA		1259	0.2%
BAC		1259	0.2%
BAX		1259	0.2%

File : cs-1

Name: Name

Value	Label	Cases	Percentage	
BBT		1259	<div></div>	0.2%
BBY		1259	<div></div>	0.2%
BDX		1259	<div></div>	0.2%
BEN		1259	<div></div>	0.2%
BF.B		1259	<div></div>	0.2%
BHF		143	<div></div>	0.0%
BHGE		152	<div></div>	0.0%
BIIB		1259	<div></div>	0.2%
BK		1259	<div></div>	0.2%
BLK		1259	<div></div>	0.2%
BLL		1259	<div></div>	0.2%
BMY		1257	<div></div>	0.2%
BRK.B		1259	<div></div>	0.2%
BSX		1259	<div></div>	0.2%
BWA		1259	<div></div>	0.2%
BXP		1259	<div></div>	0.2%
C		1259	<div></div>	0.2%
CA		1259	<div></div>	0.2%
CAG		1259	<div></div>	0.2%
CAH		1259	<div></div>	0.2%
CAT		1259	<div></div>	0.2%
CB		1259	<div></div>	0.2%
CBG		1259	<div></div>	0.2%
CBOE		1259	<div></div>	0.2%
CBS		1259	<div></div>	0.2%
CCI		1259	<div></div>	0.2%
CCL		1259	<div></div>	0.2%
CDNS		1259	<div></div>	0.2%
CELG		1259	<div></div>	0.2%
CERN		1259	<div></div>	0.2%
CF		1259	<div></div>	0.2%
CFG		850	<div></div>	0.1%
CHD		1259	<div></div>	0.2%
CHK		1259	<div></div>	0.2%
CHRW		1259	<div></div>	0.2%
CHTR		1259	<div></div>	0.2%
CI		1259	<div></div>	0.2%
CINF		1259	<div></div>	0.2%
CL		1259	<div></div>	0.2%
CLX		1259	<div></div>	0.2%
CMA		1259	<div></div>	0.2%
CMCSA		1259	<div></div>	0.2%
CME		1259	<div></div>	0.2%

File : cs-1

Name: Name

Value	Label	Cases	Percentage
CMG		1259	0.2%
CMI		1259	0.2%
CMS		1259	0.2%
CNC		1259	0.2%
CNP		1259	0.2%
COF		1259	0.2%
COG		1259	0.2%
COL		1259	0.2%
COO		1259	0.2%
COP		1259	0.2%
COST		1259	0.2%
COTY		1173	0.2%
CPB		1259	0.2%
CRM		1259	0.2%
CSCO		1259	0.2%
CSRA		561	0.1%
CSX		1259	0.2%
CTAS		1259	0.2%
CTL		1259	0.2%
CTSH		1259	0.2%
CTXS		1259	0.2%
CVS		1259	0.2%
CVX		1259	0.2%
CXO		1259	0.2%
D		1259	0.2%
DAL		1259	0.2%
DE		1259	0.2%
DFS		1259	0.2%
DG		1259	0.2%
DGX		1259	0.2%
DHI		1259	0.2%
DHR		1243	0.2%
DIS		1259	0.2%
DISCA		1259	0.2%
DISCK		1259	0.2%
DISH		1259	0.2%
DLR		1259	0.2%
DLTR		1259	0.2%
DOV		1259	0.2%
DPS		1259	0.2%
DRE		1259	0.2%
DRI		1259	0.2%
DTE		1259	0.2%

File : cs-1

Name: Name

Value	Label	Cases	Percentage
DUK		1259	<div></div> 0.2%
DVA		1259	<div></div> 0.2%
DVN		1259	<div></div> 0.2%
DWDP		109	<div></div> 0.0%
DXC		215	<div></div> 0.0%
EA		1259	<div></div> 0.2%
EBAY		1259	<div></div> 0.2%
ECL		1259	<div></div> 0.2%
ED		1259	<div></div> 0.2%
EFX		1259	<div></div> 0.2%
EIX		1259	<div></div> 0.2%
EL		1259	<div></div> 0.2%
EMN		1259	<div></div> 0.2%
EMR		1259	<div></div> 0.2%
EOG		1259	<div></div> 0.2%
EQIX		1259	<div></div> 0.2%
EQR		1259	<div></div> 0.2%
EQT		1259	<div></div> 0.2%
ES		1249	<div></div> 0.2%
ESRX		1259	<div></div> 0.2%
ESS		1259	<div></div> 0.2%
ETFC		1259	<div></div> 0.2%
ETN		1259	<div></div> 0.2%
ETR		1259	<div></div> 0.2%
EVHC		297	<div></div> 0.0%
EW		1259	<div></div> 0.2%
EXC		1259	<div></div> 0.2%
EXPD		1259	<div></div> 0.2%
EXPE		1259	<div></div> 0.2%
EXR		1259	<div></div> 0.2%
F		1259	<div></div> 0.2%
FAST		1259	<div></div> 0.2%
FB		1259	<div></div> 0.2%
FBHS		1259	<div></div> 0.2%
FCX		1259	<div></div> 0.2%
FDX		1259	<div></div> 0.2%
FE		1259	<div></div> 0.2%
FFIV		1259	<div></div> 0.2%
FIS		1259	<div></div> 0.2%
FISV		1259	<div></div> 0.2%
FITB		1259	<div></div> 0.2%
FL		1259	<div></div> 0.2%
FLIR		1259	<div></div> 0.2%

File : cs-1

Name: Name

Value	Label	Cases	Percentage
FLR		1259	0.2%
FLS		1259	0.2%
FMC		1259	0.2%
FOX		1169	0.2%
FOXA		1169	0.2%
FRT		1259	0.2%
FTI		1259	0.2%
FTV		404	0.1%
GD		1259	0.2%
GE		1259	0.2%
GGP		1259	0.2%
GILD		1259	0.2%
GIS		1259	0.2%
GLW		1259	0.2%
GM		1259	0.2%
GOOG		975	0.2%
GOOGL		1259	0.2%
GPC		1259	0.2%
GPN		1259	0.2%
GPS		1259	0.2%
GRMN		1259	0.2%
GS		1259	0.2%
GT		1259	0.2%
GWW		1259	0.2%
HAL		1259	0.2%
HAS		1259	0.2%
HBAN		1259	0.2%
HBI		1259	0.2%
HCA		1259	0.2%
HCN		1259	0.2%
HCP		1259	0.2%
HD		1259	0.2%
HES		1259	0.2%
HIG		1259	0.2%
HII		1259	0.2%
HLT		276	0.0%
HOG		1259	0.2%
HOLX		1259	0.2%
HON		1259	0.2%
HP		1259	0.2%
HPE		581	0.1%
HPQ		581	0.1%
HRB		1259	0.2%

File : cs-1

Name: Name

Value	Label	Cases	Percentage
HRL		1259	0.2%
HRS		1259	0.2%
HSIC		1259	0.2%
HST		1259	0.2%
HSY		1259	0.2%
HUM		1259	0.2%
IBM		1259	0.2%
ICE		1254	0.2%
IDXX		1259	0.2%
IFF		1259	0.2%
ILMN		1259	0.2%
INCY		1259	0.2%
INFO		917	0.1%
INTC		1259	0.2%
INTU		1259	0.2%
IP		1259	0.2%
IPG		1259	0.2%
IQV		1197	0.2%
IR		1259	0.2%
IRM		1259	0.2%
ISRG		1259	0.2%
IT		1259	0.2%
ITW		1259	0.2%
IVZ		1259	0.2%
JBHT		1259	0.2%
JCI		1259	0.2%
JEC		1259	0.2%
JNJ		1259	0.2%
JNPR		1259	0.2%
JPM		1259	0.2%
JWN		1259	0.2%
K		1259	0.2%
KEY		1259	0.2%
KHC		655	0.1%
KIM		1259	0.2%
KLAC		1259	0.2%
KMB		1259	0.2%
KMI		1259	0.2%
KMX		1259	0.2%
KO		1259	0.2%
KORS		1259	0.2%
KR		1259	0.2%
KSS		1259	0.2%

File : cs-1

Name: Name

Value	Label	Cases	Percentage	
KSU		1259		0.2%
L		1259		0.2%
LB		1259		0.2%
LEG		1259		0.2%
LEN		1259		0.2%
LH		1259		0.2%
LKQ		1259		0.2%
LLL		1259		0.2%
LLY		1259		0.2%
LMT		1259		0.2%
LNC		1259		0.2%
LNT		1259		0.2%
LOW		1259		0.2%
LRCX		1259		0.2%
LUK		1259		0.2%
LUV		1259		0.2%
LYB		1259		0.2%
M		1259		0.2%
MA		1259		0.2%
MAA		1259		0.2%
MAC		1259		0.2%
MAR		1259		0.2%
MAS		1259		0.2%
MAT		1259		0.2%
MCD		1259		0.2%
MCHP		1259		0.2%
MCK		1259		0.2%
MCO		1259		0.2%
MDLZ		1259		0.2%
MDT		1259		0.2%
MET		1259		0.2%
MGM		1259		0.2%
MHK		1259		0.2%
MKC		1259		0.2%
MLM		1259		0.2%
MMC		1259		0.2%
MMM		1259		0.2%
MNST		1259		0.2%
MO		1259		0.2%
MON		1259		0.2%
MOS		1259		0.2%
MPC		1259		0.2%
MRK		1259		0.2%

File : cs-1

Name: Name

Value	Label	Cases	Percentage
MRO		1259	0.2%
MS		1259	0.2%
MSFT		1259	0.2%
MSI		1259	0.2%
MTB		1259	0.2%
MTD		1259	0.2%
MU		1259	0.2%
MYL		1259	0.2%
NAVI		960	0.2%
NBL		1259	0.2%
NCLH		1259	0.2%
NDAQ		1259	0.2%
NEE		1259	0.2%
NEM		1259	0.2%
NFLX		1259	0.2%
NFX		1259	0.2%
NI		1259	0.2%
NKE		1259	0.2%
NLSN		1259	0.2%
NOC		1259	0.2%
NOV		1259	0.2%
NRG		1259	0.2%
NSC		1259	0.2%
NTAP		1259	0.2%
NTRS		1259	0.2%
NUE		1259	0.2%
NVDA		1259	0.2%
NWL		1259	0.2%
NWS		1169	0.2%
NWSA		1169	0.2%
O		1243	0.2%
OKE		1259	0.2%
OMC		1259	0.2%
ORCL		1257	0.2%
ORLY		1259	0.2%
OXY		1259	0.2%
PAYX		1259	0.2%
PBCT		1259	0.2%
PCAR		1259	0.2%
PCG		1259	0.2%
PCLN		1259	0.2%
PDCO		1259	0.2%
PEG		1259	0.2%

File : cs-1

Name: Name

Value	Label	Cases	Percentage	
PEP		1259		0.2%
PFE		1259		0.2%
PFG		1259		0.2%
PG		1259		0.2%
PGR		1259		0.2%
PH		1259		0.2%
PHM		1259		0.2%
PKG		1259		0.2%
PKI		1259		0.2%
PLD		1259		0.2%
PM		1259		0.2%
PNC		1259		0.2%
PNR		1259		0.2%
PNW		1259		0.2%
PPG		1259		0.2%
PPL		1259		0.2%
PRGO		1259		0.2%
PRU		1259		0.2%
PSA		1259		0.2%
PSX		1259		0.2%
PVH		1259		0.2%
PWR		1259		0.2%
PX		1259		0.2%
PXD		1259		0.2%
PYPL		655	0.1%	
QCOM		1259		0.2%
QRVO		781	0.1%	
RCL		1259		0.2%
RE		1259		0.2%
REG		1259		0.2%
REGN		1259		0.2%
RF		1259		0.2%
RHI		1259		0.2%
RHT		1259		0.2%
RJF		1259		0.2%
RL		1259		0.2%
RMD		1259		0.2%
ROK		1259		0.2%
ROP		1259		0.2%
ROST		1259		0.2%
RRC		1259		0.2%
RSG		1259		0.2%
RTN		1259		0.2%

File : cs-1

Name: Name

Value	Label	Cases	Percentage	
SBAC		1259		0.2%
SBUX		1259		0.2%
SCG		1259		0.2%
SCHW		1259		0.2%
SEE		1259		0.2%
SHW		1259		0.2%
SIG		1259		0.2%
SJM		1259		0.2%
SLB		1259		0.2%
SLG		1259		0.2%
SNA		1259		0.2%
SNI		1259		0.2%
SNPS		1259		0.2%
SO		1259		0.2%
SPG		1259		0.2%
SPGI		1259		0.2%
SRCL		1259		0.2%
SRE		1259		0.2%
STI		1259		0.2%
STT		1259		0.2%
STX		1259		0.2%
STZ		1259		0.2%
SWK		1259		0.2%
SWKS		1259		0.2%
SYF		888		0.1%
SYK		1259		0.2%
SYMC		1259		0.2%
SYU		1259		0.2%
T		1259		0.2%
TAP		1259		0.2%
TDG		1259		0.2%
TEL		1259		0.2%
TGT		1259		0.2%
TIF		1259		0.2%
TJX		1259		0.2%
TMK		1259		0.2%
TMO		1259		0.2%
TPR		1259		0.2%
TRIP		1259		0.2%
TROW		1259		0.2%
TRV		1259		0.2%
TSCO		1259		0.2%
TSN		1259		0.2%

File : cs-1

Name: Name

Value	Label	Cases	Percentage
TSS		1259	0.2%
TWX		1259	0.2%
TXN		1259	0.2%
TXT		1259	0.2%
UA		464	0.1%
UAA		1259	0.2%
UAL		1259	0.2%
UDR		1259	0.2%
UHS		1259	0.2%
ULTA		1259	0.2%
UNH		1259	0.2%
UNM		1259	0.2%
UNP		1259	0.2%
UPS		1259	0.2%
URI		1259	0.2%
USB		1259	0.2%
UTX		1259	0.2%
V		1259	0.2%
VAR		1259	0.2%
VFC		1259	0.2%
VIAB		1259	0.2%
VLO		1259	0.2%
VMC		1259	0.2%
VNO		1259	0.2%
VRSK		1259	0.2%
VRSN		1259	0.2%
VRTX		1259	0.2%
VTR		1259	0.2%
VZ		1259	0.2%
WAT		1259	0.2%
WBA		1259	0.2%
WDC		1259	0.2%
WEC		1259	0.2%
WFC		1259	0.2%
WHR		1259	0.2%
WLTW		528	0.1%
WM		1259	0.2%
WMB		1259	0.2%
WMT		1259	0.2%
WRK		662	0.1%
WU		1259	0.2%
WY		1259	0.2%
WYN		1259	0.2%

File : cs-1

Name: Name

Value	Label	Cases	Percentage	
WYNN		1259		0.2%
XEC		1259		0.2%
XEL		1259		0.2%
XL		1259		0.2%
XLNX		1259		0.2%
XOM		1259		0.2%
XRAY		1259		0.2%
XRX		1259		0.2%
XYL		1259		0.2%
YUM		1259		0.2%
ZBH		1259		0.2%
ZION		1259		0.2%
ZTS		1259		0.2%

Warning: these figures indicate the number of cases found in the data file. They cannot be interpreted as summary statistics of the population of interest.