

WRDS_database

January 11, 2022

1 WRDS database use case

1. [Package](#)
2. [Connection](#)
3. [Database](#)
4. [Tables](#)
5. Stock Data
6. [Example](#)

1.1 Package

1. Required package from [WRDS](#)

```
[ ]: import wrds
```

1.2 Connection

2. Make connection to WRDS database via the API

In connection function `wrds_username` is the login username on WRDS website. The password will be prompted on the first execution. By running `create_pgpass_file()`, the input of password can be skipped for next time.

```
[ ]: conn = wrds.Connection(wrds_username='chenglu')
```

Loading library list...
Done

1.3 Database

3. Check all available [libraries](#) in WRDS

There 252 libraries (database) available in WRDS (2022/01/11). In this tutorial we consider the [CRSP](#)(Center for Research in Security Prices). A brief view of part of CRSP database is shown below.

Product	Description
crsp_a_ccm	CRSP/Compustat Merged (Annual)
crsp_a_indexes	CRSP Indexes (Annual)
crsp_a_stock	CRSP Stock (Annual)

Product	Description
crsp_a_stock10	CRSP10 is a variation of the CRSP Stock database and holds 10 years of monthly history.
crsp_a_stock62	CRSP Stock 1962 (Annual)
crsp_a_treasuries	CRSP Treasuries (Annual)
crsp_a_ziman	CRSP Ziman Real Estates (Annual)

```
[ ]: library_list = conn.list_libraries()
print(library_list[0:5])
```

```
['bank', 'compbd', 'compgd', 'compnad', 'compsegd']
```

1.4 Tables

4. Check available tables within CRSP

CRSP database focus on the stock data on the U.S. market. There are 42 tables within CRSP databse. (2022/01/11) A sample description of the sub-table is listed here. More details can be found in [CRSP website](#)

Table	Description
dse	Daily Stock - Events
dseall	None
dsedelist	CRSP Daily Stock Event - Delisting
dsedist	CRSP Daily Stock Event -Distribution
dseexchdates	None
dsenames	CRSP Daily Stock Event - Name History
dsenasdin	CRSP Daily Stock Event - NASDAQ Information
dseshares	CRSP Daily Stock Event - Shares Outstanding
dsf	Daily Stock - Securities
dsfhdr	Daily Stock - Header Information with Date Ranges
dsi	Stock - Market Indexes Daily NYSE/AMEX/NASDAQ/ARCA

```
[ ]: table_list = conn.list_tables(library='crsp_a_stock')
print(table_list[0: 10])
```

```
['dse', 'dseall', 'dsedelist', 'dsedist', 'dseexchdates', 'dsenames',
'dsenasdin', 'dseshares', 'dsf', 'dsfhdr']
```

1.5 Stock Data

5. Daily stock data

The main table we look into is **dse**, which stands for daily stock events. It contains the stock price, stock return and etc. A brief description is shown below. More details in [CRSP website](#).

Variable Name	Type	Length	Description
ask	double	53	Ask
askhi	double	53	Ask or High Price
bid	double	53	Bid
bidlo	double	53	Bid or Low Price
cfacpr	double	53	Cumulative Factor to Adjust Prices
cfacshr	double	53	Cumulative Factor to Adjust Shares/Vol
cusip	string	8	CUSIP Header
date	date		Date of Observation
hexcd	double	53	Exchange Code Header
hsiccd	double	53	Standard Industrial Classification Code
issuno	double	53	Nasdaq Issue Number
numtrd	double	53	Number of Trades
openprc	double	53	Price Alternate
permco	double	53	PERMCO
permno	double	53	PERMNO
prc	double	53	Price or Bid/Ask Average
ret	double	53	Returns
retx	double	53	Returns without Dividends
shrout	double	53	Shares Outstanding
vol	double	53	Volume

```
[ ]: data = conn.get_table(library='crsp', table='dse', obs=3)
data
```

```
[ ]:      event      date  hsicmg  hsicig      comnam      cusip  \
0  NASDIN  1986-01-06   39.0   399.0      None  68391610
1  NAMES  1986-01-07   39.0   399.0  OPTIMUM MANUFACTURING INC  68391610
2  NASDIN  1986-01-07   39.0   399.0      None  68391610

      dclrdt dlamt dlpdt dlstcd  ...  dlretx  dlprc dlret shrout shrenddt trtscd  \
0  None  None  None  None  ...  None  None  None  None  None  4.0
1  None  None  None  None  ...  None  None  None  None  None  NaN
2  None  None  None  None  ...  None  None  None  None  None  1.0

      trtsenddt nmsind mmcnt  nsdinx
0  1986-01-06   1.0   0.0   1.0
1      None   NaN   NaN   NaN
2  1986-01-09   1.0   9.0   2.0
```

[3 rows x 50 columns]

6. Permno code

The **permno** code is an identifier in CRSP database given to each stock. One can look up related information and data about one stock using its **permno** code. One company's name and its ticker name can change over time. The table **dse** contains the history of this kind of information

for each stock.

```
[ ]: data = conn.get_table(library='crsp', table='dsenames', obs=3)
data.head()
```

```
[ ]:      permno      namedt      nameendt  shrcd  exchcd  siccd  ncusip ticker \
0  10000.0  1986-01-07  1986-12-03   10.0    3.0  3990.0  68391610  OMFGA
1  10000.0  1986-12-04  1987-03-09   10.0    3.0  3990.0  68391610  OMFGA
2  10000.0  1987-03-10  1987-06-11   10.0    3.0  3990.0  68391610  OMFGA

      comnam shrcsls  ... naics primexch trdstat secstat \
0  OPTIMUM MANUFACTURING INC    A  ... None    Q    A    R
1  OPTIMUM MANUFACTURING INC    A  ... None    Q    A    R
2  OPTIMUM MANUFACTURING INC    A  ... None    Q    A    R

      permco      compno      issuno  hexcd  hsiccd      cusip
0  7952.0  60007905.0  10396.0    3.0  3990.0  68391610
1  7952.0  60007905.0  10396.0    3.0  3990.0  68391610
2  7952.0  60007905.0  10396.0    3.0  3990.0  68391610
```

[3 rows x 21 columns]

7. stock data download

It's a two-step procedure to download the stock daily data from CRSP. First, a **permno** code is obtained for a given stock symbol name. Second, we can use the **permno** code to query the **dse** table to download daily data.

```
[ ]: query = "\
SELECT * \
FROM crsp.dsenames \
WHERE ticker = 'AMZN'\
"

ticker_history = conn.raw_sql(query)
ticker_history
```

```
[ ]:      permno      namedt      nameendt  shrcd  exchcd  siccd  ncusip ticker \
0  84788.0  1997-05-15  2004-06-09   11.0    3.0  7370.0  02313510  AMZN
1  84788.0  2004-06-10  2014-01-14   11.0    3.0  7370.0  02313510  AMZN
2  84788.0  2014-01-15  2020-03-18   11.0    3.0  7370.0  02313510  AMZN
3  84788.0  2020-03-19  2020-12-31   11.0    3.0  7370.0  02313510  AMZN

      comnam shrcsls  ... naics primexch trdstat secstat  permco \
0  AMAZON COM INC  None  ... None    Q    A    R  15473.0
1  AMAZON COM INC  None  ... 454110    Q    A    R  15473.0
2  AMAZON COM INC  None  ... 454113    Q    A    R  15473.0
3  AMAZON COM INC  None  ... 454110    Q    A    R  15473.0
```

	compno	issuno	hexcd	hsiccd	cusip
0	60015310.0	20733.0	3.0	7370.0	02313510
1	60015310.0	20733.0	3.0	7370.0	02313510
2	60015310.0	20733.0	3.0	7370.0	02313510
3	60015310.0	20733.0	3.0	7370.0	02313510

[4 rows x 21 columns]

```
[ ]: query = "\
SELECT *\
FROM crsp.dsf \
WHERE permno = 14593 \
AND date BETWEEN '1996-01-01' AND '2020-01-01'"

stock_daily_data = conn.raw_sql(query)
stock_daily_data.head()
```

```
[ ]:      cusip  permno  permco  issuno  hexcd  hsiccd      date  bidlo  \
0  03783310  14593.0    7.0    8.0    3.0  3571.0  1996-01-02  31.750
1  03783310  14593.0    7.0    8.0    3.0  3571.0  1996-01-03  31.875
2  03783310  14593.0    7.0    8.0    3.0  3571.0  1996-01-04  31.375
3  03783310  14593.0    7.0    8.0    3.0  3571.0  1996-01-05  31.375
4  03783310  14593.0    7.0    8.0    3.0  3571.0  1996-01-08  34.000

      askhi    prc      vol      ret    bid    ask    shrout  cfacpr  \
0  32.250  32.1250  1249047.0  0.007843  32.000  32.125  123118.0  112.0
1  32.875  32.1250  3848106.0  0.000000  32.125  32.250  123118.0  112.0
2  32.375  31.5625  2686841.0 -0.017510  31.500  31.625  123118.0  112.0
3  34.250  34.2500  3989514.0  0.085149  34.125  34.250  123118.0  112.0
4  35.500  34.6250  1086787.0  0.010949  34.375  34.625  123118.0  112.0

      cfacshr  openprc  numtrd      retx
0    112.0    32.250  1423.0  0.007843
1    112.0    32.000  2281.0  0.000000
2    112.0    32.375  2208.0 -0.017510
3    112.0    31.625  3408.0  0.085149
4    112.0    34.500  1162.0  0.010949
```

1.6 Example

Following example shows a concrete usage of CRSP.

Download Dow 30 constitutes daily stock data as an example of WRDS API

1. Full list of Dow 30
2. Ticker to PERMNO map
3. Pannel data

4. Long table to wide table

```
[ ]: import pandas as pd
import yahoo_fin.stock_info as si
```

1.6.1 1. Full list of Dow 30

Download Dow 30 stocks ticker names from wiki page using package `yahoo_fin`. It's different from the famous `yfinance`. The documentation of `yahoo_fin` can be found [here](#).

```
[ ]: dow_tickers_df = si.tickers_dow(True)
dow_tickers_list = si.tickers_dow()
dow_tickers_df.head()
```

```
[ ]:      Company Exchange Symbol      Industry Date added \
0          3M      NYSE   MMM      Conglomerate 1976-08-09
1 American Express      NYSE   AXP      Financial services 1982-08-30
2          Amgen  NASDAQ  AMGN      Biopharmaceutical 2020-08-31
3          Apple  NASDAQ  AAPL      Information technology 2015-03-19
4          Boeing      NYSE   BA      Aerospace and defense 1987-03-12
```

```
      Notes Index weighting
0 As Minnesota Mining and Manufacturing      3.38%
1      NaN      3.29%
2      NaN      3.84%
3      NaN      2.76%
4      NaN      4.01%
```

1.6.2 2. Ticker to PERMNO map

```
[ ]: dow_tickers_tuple_str = str(tuple(dow_tickers_list))

query = "\
SELECT *\
FROM crsp.dsenames \
WHERE ticker in {} \
AND nameenddt = '2020-12-31' \
".format(dow_tickers_tuple_str)

ticker_history = conn.raw_sql(query)
ticker_history.head()
```

```
[ ]:      permno      namedt      nameenddt      shrccd      exchcd      siccd      ncusip      ticker \
0 14593.0 2017-12-28 2020-12-31 11.0 3.0 3571.0 03783310 AAPL
1 14008.0 2019-09-12 2020-12-31 11.0 3.0 2830.0 03116210 AMGN
2 59176.0 2020-05-01 2020-12-31 11.0 1.0 6029.0 02581610 AXP
3 19561.0 2004-06-10 2020-12-31 11.0 1.0 3721.0 09702310 BA
```

4	18542.0	2017-01-26	2020-12-31	11.0	1.0	3531.0	14912310	CAT
---	---------	------------	------------	------	-----	--------	----------	-----

	comnam	shrcls	...	naics	primexch	trdstat	secstat	permco	\
0	APPLE INC	None	...	334220	Q	A	R	7.0	
1	AMGEN INC	None	...	325414	Q	A	R	216.0	
2	AMERICAN EXPRESS CO	None	...	522110	N	A	R	90.0	
3	BOEING CO	None	...	336411	N	A	R	20315.0	
4	CATERPILLAR INC	None	...	333120	N	A	R	20408.0	

	compno	issuno	hexcd	hsiccd	cusip
0	60000006.0	8.0	3.0	3571.0	03783310
1	60000215.0	298.0	3.0	2830.0	03116210
2	60000089.0	126.0	1.0	6029.0	02581610
3	0.0	0.0	1.0	3721.0	09702310
4	0.0	0.0	1.0	3531.0	14912310

[5 rows x 21 columns]

create a map to convert ticker names to permno code

```
[ ]: ticker_permno_df = ticker_history[['permno', 'ticker']]
      ticker_to_permno_map = ticker_permno_df.set_index('ticker')
      permno_to_ticker_map = ticker_permno_df.set_index('permno')

[ ]: not_complete_permno_list = [18428, 76076, 86868, 90215, 92611]
      permno_to_ticker_map.loc[not_complete_permno_list]
```

```
[ ]:      ticker
      permno
18428      DOW
76076      CSCO
86868       GS
90215      CRM
92611       V
```

1.6.3 3. Pannel data of Dow 30

Query dse to get Dow 30 stocks' daily data from 1990 to 2019. Download data is a pannel data. One complete time series for one stock followed by another one.

```
[ ]: dow_permno_tuple_str = str(tuple(ticker_to_permno_map.loc[dow_tickers_list].
    ↪values.reshape((1, -1))[0].tolist()))

      query = "\
      SELECT *\
      FROM crsp.dsf \
      WHERE permno in {} \
```

```
AND date BETWEEN '1990-01-01' AND '2020-01-01'\
".format(dow_permno_tuple_str)
```

```
dow_df = conn.raw_sql(query)
dow_df.head()
```

```
[ ]:      cusip    permno  permco   issuno  hexcd  hsiccd      date  bidlo  \
0  59491810  10107.0  8048.0  10539.0    3.0  7370.0  1990-01-02  86.125
1  59491810  10107.0  8048.0  10539.0    3.0  7370.0  1990-01-03  88.500
2  59491810  10107.0  8048.0  10539.0    3.0  7370.0  1990-01-04  88.750
3  59491810  10107.0  8048.0  10539.0    3.0  7370.0  1990-01-05  89.500
4  59491810  10107.0  8048.0  10539.0    3.0  7370.0  1990-01-08  88.500

      askhi    prc      vol      ret    bid    ask  shrout  cfacpr  cfacshr  \
0  88.75  88.750  369793.0  0.020115  88.25  88.75  55676.0  144.0  144.0
1  90.25  89.250  791777.0  0.005634  89.25  89.75  55676.0  144.0  144.0
2  92.00  91.875  874533.0  0.029412  91.75  92.00  55676.0  144.0  144.0
3  92.00  89.625  484027.0 -0.024490  89.50  90.00  55676.0  144.0  144.0
4  91.00  91.000  410461.0  0.015342  90.75  91.00  55676.0  144.0  144.0

      openprc  numtrd      retx
0         NaN   409.0  0.020115
1         NaN   671.0  0.005634
2         NaN   614.0  0.029412
3         NaN   533.0 -0.024490
4         NaN   337.0  0.015342
```

1.6.4 4. Dow 30 ret dataframe

Convert pannel data to dataframe, which use timestamp as index and permno code as column names.

```
[ ]: dow_df = dow_df.set_index(['permno', 'date'])
dow_ret_df = dow_df['ret']
dow_ret_df = dow_ret_df.unstack().T
dow_ret_df.head()
```

```
[ ]: permno      10107.0   10145.0   11308.0   12490.0   14008.0   14541.0  \
date
1990-01-02  0.020115  0.014337  0.009709  0.041169  0.051020  0.020295
1990-01-03  0.005634  0.003534 -0.014423  0.008929  0.024272 -0.016275
1990-01-04  0.029412  0.031690 -0.004878  0.011378  0.018957 -0.012868
1990-01-05 -0.024490  0.013652 -0.008170 -0.002500  0.013953 -0.014898
1990-01-08  0.015342  0.000000  0.021417  0.006266 -0.002294  0.009452

permno      14593.0   18163.0  18428.0   18542.0  ...   59176.0   59328.0  \
date
...
```


1990-01-02	0.056738	0.003559	NaN	0.017279	...	0.000000	0.043478
1990-01-03	0.006711	-0.001773	NaN	0.010616	...	0.010753	-0.027778
1990-01-04	0.003333	-0.019538	NaN	0.004202	...	-0.010638	0.021429
1990-01-05	0.003322	-0.014493	NaN	-0.006276	...	-0.010753	-0.006993
1990-01-08	0.006623	0.014706	NaN	-0.006316	...	-0.007246	0.014085

permno	59459.0	65875.0	66181.0	76076.0	86868.0	90215.0	92611.0	\
date								
1990-01-02	0.014799	0.024719	0.030717	NaN	NaN	NaN	NaN	
1990-01-03	0.006250	-0.002193	0.003311	NaN	NaN	NaN	NaN	
1990-01-04	0.018634	-0.028791	0.006601	NaN	NaN	NaN	NaN	
1990-01-05	-0.004065	-0.022857	-0.013115	NaN	NaN	NaN	NaN	
1990-01-08	0.004082	0.008187	-0.009967	NaN	NaN	NaN	NaN	

permno	92655.0
date	
1990-01-02	0.010309
1990-01-03	-0.020408
1990-01-04	-0.010417
1990-01-05	-0.042105
1990-01-08	-0.021978

[5 rows x 30 columns]

1.6.5 5. Merge Dow with given ret data

```
[ ]: ret_data = pd.read_csv('/Users/cheng/Google Drive/PhD/Research/Portfolio_
    ↪Selection via TBN/data/Data/permno_ret.csv', parse_dates = True, index_col=0)
ret_data.columns = ret_data.columns.astype(int)

[ ]: ticker_intersection_set = set(ret_data.columns).intersection(set(dow_ret_df.
    ↪columns))
ticker_intersection_list = list(ticker_intersection_set)

dow_ret_df_nonlap = dow_ret_df.drop(ticker_intersection_list, axis=1)

data = pd.concat([ret_data, dow_ret_df_nonlap], axis=1)

[ ]: data.to_csv('/Users/cheng/Google Drive/PhD/Research/Portfolio Selection via TBN/
    ↪data/Data/stock_return_crsp_159.csv')
```