Homework Sheet

CEBD 1160 – Intro to Big Data Technology – Winter 2020 – Jan 18 to Mar 21

Student Name: Lan (Rebecca) Xu

Class 1: [Intro 2 Course & Big Data](https://cce-bigdataintro-1160.github.io/winter2020-site/1-intro.html)

**Journal:**

Submitted

**Basic:**

Slack is set up

My GitHub page link: <https://github.com/mtl-lan>

**Advanced:**

a. Chosen Dataset name: US Accidents (3.0 million records) – A Countrywide Traffic Accident Dataset (2016-2019)16

Dataset link: <https://www.kaggle.com/sobhanmoosavi/us-accidents>

b. The characteristics of the chosen dataset:

- Dataset format: csv

- Dataset size: 1 GB

- Dataset is structured. (49 headers including accident location, weather condition and accident time etc.)

- Dataset frequency: keep updating. (First released in May 2019, and it is updated in Jan 2020)

- Data reliability: This data has been collected in real-time, using multiple Traffic APIs from two sources: MapQuest and Bing.

- Data completeness: not 100% complete since there is no data collected by *MapQuest* from Feb 2016 to August 2017: AL, AR, AZ, CO, ID, KS, KY, LA, ME, MN, MS, MT, NC, ND, NH, NM, NV, OK, OR, SD, TN, UT, VT, WI, and WY.

c. Research question

As a public traffic governor, how to decrease the car accident in US? Find out the top N feasible solutions.

**Reach**

- My home computer is Linux Ubuntu

cat /proc/version

Linux version 4.18.0-25-generic (buildd@lgw01-amd64-033) (gcc version 7.4.0 (Ubuntu 7.4.0-1ubuntu1~18.04.1)) #26~18.04.1-Ubuntu SMP Thu Jun 27 07:28:31 UTC 2019

- Open the Terminal

ctrl+alt+t

- Running the following commands: ls/date/pwd/whoami/env, the outputs are:

rebecca@Torex-Power: ~$ ls

anaconda3 cudnn\_samples\_v7 Documents examples.desktop lanlan myfile Pictures PycharmProjects Templates

bin Desktop Downloads get-docker.sh Music nvvp\_workspace Public snap Videos

rebecca@Torex-Power: ~$ date

Tue Jan 21 11:47:23 EST 2020

rebecca@Torex-Power: ~$ pwd

/home/rebecca

rebecca@Torex-Power:~$ whoami

rebecca

rebecca@Torex-Power: ~$ env

CLUTTER\_IM\_MODULE=xim

CONDA\_SHLVL=0

LD\_LIBRARY\_PATH=:/usr/local/cuda/lib64:/usr/local/cuda/extras/CUPTI/lib64

LS\_COLORS=rs=0:di=01;34:ln=01;36:mh=00:pi=40;33:so=01;35:do=01;35:bd=40;33;01:cd=40;33;01:or=40;31;01:mi=00:su=37;41:sg=30;43:ca=30;41:tw=30;42:ow=34;42:st=37;44:ex=01;32:\*.tar=01;31:\*.tgz=01;31:\*.arc=01;31:\*.arj=01;31:\*.taz=01;31:\*.lha=01;31:\*.lz4=01;31:\*.lzh=01;31:\*.lzma=01;31:\*.tlz=01;31:\*.txz=01;31:\*.tzo=01;31:\*.t7z=01;31:\*.zip=01;31:\*.z=01;31:\*.Z=01;31:\*.dz=01;31:\*.gz=01;31:\*.lrz=01;31:\*.lz=01;31:\*.lzo=01;31:\*.xz=01;31:\*.zst=01;31:\*.tzst=01;31:\*.bz2=01;31:\*.bz=01;31:\*.tbz=01;31:\*.tbz2=01;31:\*.tz=01;31:\*.deb=01;31:\*.rpm=01;31:\*.jar=01;31:\*.war=01;31:\*.ear=01;31:\*.sar=01;31:\*.rar=01;31:\*.alz=01;31:\*.ace=01;31:\*.zoo=01;31:\*.cpio=01;31:\*.7z=01;31:\*.rz=01;31:\*.cab=01;31:\*.wim=01;31:\*.swm=01;31:\*.dwm=01;31:\*.esd=01;31:\*.jpg=01;35:\*.jpeg=01;35:\*.mjpg=01;35:\*.mjpeg=01;35:\*.gif=01;35:\*.bmp=01;35:\*.pbm=01;35:\*.pgm=01;35:\*.ppm=01;35:\*.tga=01;35:\*.xbm=01;35:\*.xpm=01;35:\*.tif=01;35:\*.tiff=01;35:\*.png=01;35:\*.svg=01;35:\*.svgz=01;35:\*.mng=01;35:\*.pcx=01;35:\*.mov=01;35:\*.mpg=01;35:\*.mpeg=01;35:\*.m2v=01;35:\*.mkv=01;35:\*.webm=01;35:\*.ogm=01;35:\*.mp4=01;35:\*.m4v=01;35:\*.mp4v=01;35:\*.vob=01;35:\*.qt=01;35:\*.nuv=01;35:\*.wmv=01;35:\*.asf=01;35:\*.rm=01;35:\*.rmvb=01;35:\*.flc=01;35:\*.avi=01;35:\*.fli=01;35:\*.flv=01;35:\*.gl=01;35:\*.dl=01;35:\*.xcf=01;35:\*.xwd=01;35:\*.yuv=01;35:\*.cgm=01;35:\*.emf=01;35:\*.ogv=01;35:\*.ogx=01;35:\*.aac=00;36:\*.au=00;36:\*.flac=00;36:\*.m4a=00;36:\*.mid=00;36:\*.midi=00;36:\*.mka=00;36:\*.mp3=00;36:\*.mpc=00;36:\*.ogg=00;36:\*.ra=00;36:\*.wav=00;36:\*.oga=00;36:\*.opus=00;36:\*.spx=00;36:\*.xspf=00;36:

CONDA\_EXE=/home/rebecca/anaconda3/bin/conda

LESSCLOSE=/usr/bin/lesspipe %s %s

XDG\_MENU\_PREFIX=gnome-

LANG=en\_CA.UTF-8

DISPLAY=:1

GNOME\_SHELL\_SESSION\_MODE=ubuntu

COLORTERM=truecolor

USERNAME=rebecca

XDG\_VTNR=2

SSH\_AUTH\_SOCK=/run/user/1000/keyring/ssh

\_CE\_M=

XDG\_SESSION\_ID=3

USER=rebecca

DESKTOP\_SESSION=ubuntu

QT4\_IM\_MODULE=xim

TEXTDOMAINDIR=/usr/share/locale/

GNOME\_TERMINAL\_SCREEN=/org/gnome/Terminal/screen/a64085ed\_2bfa\_4ad5\_8c7d\_89ccd62336df

PWD=/home/rebecca

HOME=/home/rebecca

CONDA\_PYTHON\_EXE=/home/rebecca/anaconda3/bin/python

TEXTDOMAIN=im-config

SSH\_AGENT\_PID=2296

CUDA\_HOME=/usr/local/cuda

QT\_ACCESSIBILITY=1

XDG\_SESSION\_TYPE=x11

XDG\_DATA\_DIRS=/usr/share/ubuntu:/usr/local/share:/usr/share:/var/lib/snapd/desktop

\_CE\_CONDA=

XDG\_SESSION\_DESKTOP=ubuntu

GJS\_DEBUG\_OUTPUT=stderr

GTK\_MODULES=gail:atk-bridge

WINDOWPATH=2

TERM=xterm-256color

SHELL=/bin/bash

VTE\_VERSION=5202

QT\_IM\_MODULE=ibus

XMODIFIERS=@im=ibus

IM\_CONFIG\_PHASE=2

XDG\_CURRENT\_DESKTOP=ubuntu:GNOME

GPG\_AGENT\_INFO=/run/user/1000/gnupg/S.gpg-agent:0:1

GNOME\_TERMINAL\_SERVICE=:1.115

XDG\_SEAT=seat0

SHLVL=1

LANGUAGE=en\_CA:en

GDMSESSION=ubuntu

GNOME\_DESKTOP\_SESSION\_ID=this-is-deprecated

LOGNAME=rebecca

DBUS\_SESSION\_BUS\_ADDRESS=unix:path=/run/user/1000/bus

XDG\_RUNTIME\_DIR=/run/user/1000

XAUTHORITY=/run/user/1000/gdm/Xauthority

XDG\_CONFIG\_DIRS=/etc/xdg/xdg-ubuntu:/etc/xdg

PATH=/home/rebecca/bin:/usr/local/cuda/bin:/home/rebecca/anaconda3/bin:/home/rebecca/anaconda3/condabin:/home/rebecca/bin:/usr/local/sbin:/usr/local/bin:/usr/sbin:/usr/bin:/sbin:/bin:/usr/games:/usr/local/games:/snap/bin

GJS\_DEBUG\_TOPICS=JS ERROR;JS LOG

SESSION\_MANAGER=local/Torex-Power:@/tmp/.ICE-unix/2219,unix/Torex-Power:/tmp/.ICE-unix/2219

LESSOPEN=| /usr/bin/lesspipe %s

GTK\_IM\_MODULE=ibus

\_=/usr/bin/env

**Optional homework (no grading)**

**- What is an operation system shell. What’s its purpose?**

In computing, a shell is a user interface for access to an operating system's services. In general, operating system shells use either a command-line interface (CLI) or graphical user interface (GUI), depending on a computer's role and particular operation.

From: <https://en.wikipedia.org/wiki/Shell_(computing>[)](https://en.wikipedia.org/wiki/Shell_(computing))

# Why is a shell needed in an operating system? Could an operating system work without it?

If you start using the command-line as your normal interface for the computer, you will discover another advantage: there is no better interface to do text-orientated work on a computer than a text-orientated interface.

And most of your work is text- and not graphical orientated, I guess. Writing documents, emails, chats, programming, system operations and maintenance as well as configuration work are all best done on a shell.

From:<https://www.quora.com/Why-is-a-shell-needed-in-an-operating-system-Could-an-operating-system-work-without-it>

**- What is a versioning control system. What's its purpose?**

Version control is a system that records changes to a file or set of files over time so that you can recall specific versions later.

If you are a graphic or web designer and want to keep every version of an image or layout (which you would most certainly want to), a Version Control System (VCS) is a very wise thing to use. It allows you to revert selected files back to a previous state, revert the entire project back to a previous state, compare changes over time, see who last modified something that might be causing a problem, who introduced an issue and when, and more. Using a VCS also generally means that if you screw things up or lose files, you can easily recover. In addition, you get all this for very little overhead.

From: <https://git-scm.com/book/en/v2/Getting-Started-About-Version-Control>

**Home Environment preparation:**

- Git installed under Linux and set up a connection to GitHub.

- Docker installed under Linux.

- Pyhon3 installed and connected to VSCode.