



TECNOLÓGICO NACIONAL DE MÉXICO

INSTITUTO TECNOLÓGICO DE TIJUANA

SUBDIRECCIÓN ACADÉMICA

DEPARTAMENTO DE SISTEMAS Y COMPUTACIÓN

PERIODO: Agosto – Diciembre 2018

Ing. Sistemas Computacionales

Datos Masivos.  
Serie: BDD-1704 SC9A

Evaluacion

Alumna:  
Cruz Espinoza Vianey Kareli   
14212013

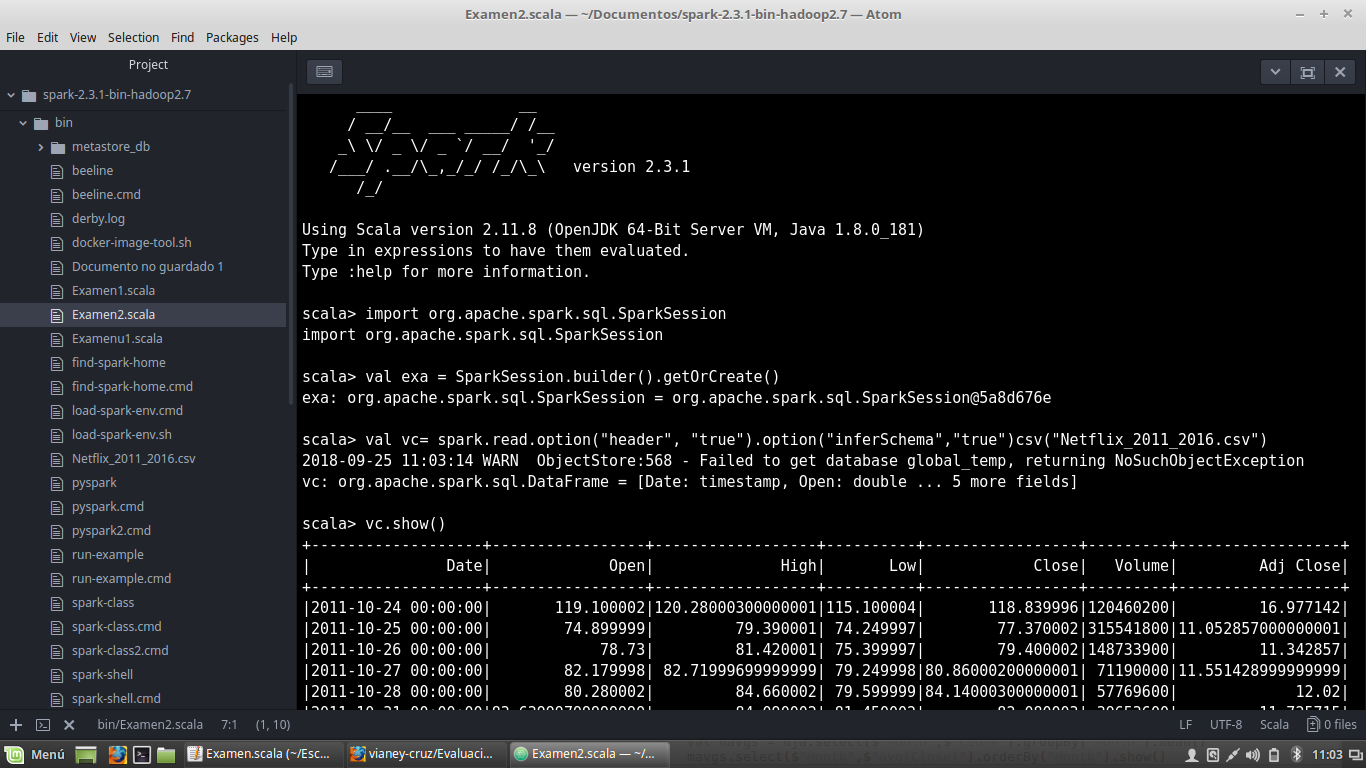
Catedrático:

Romero Hernandez Jose Christian.

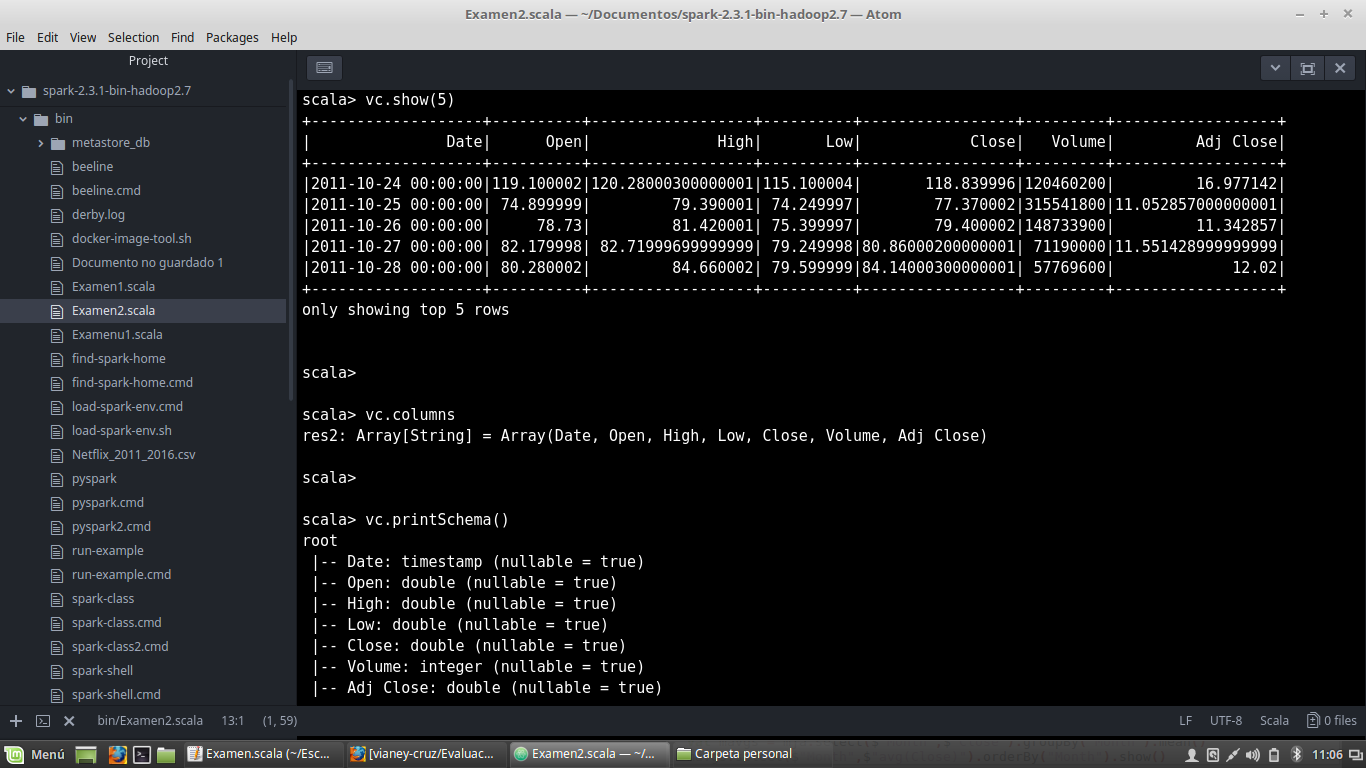
Tijuana, Baja California a 20 de Septiembre del 2018

## Comienza una simple Spark Session

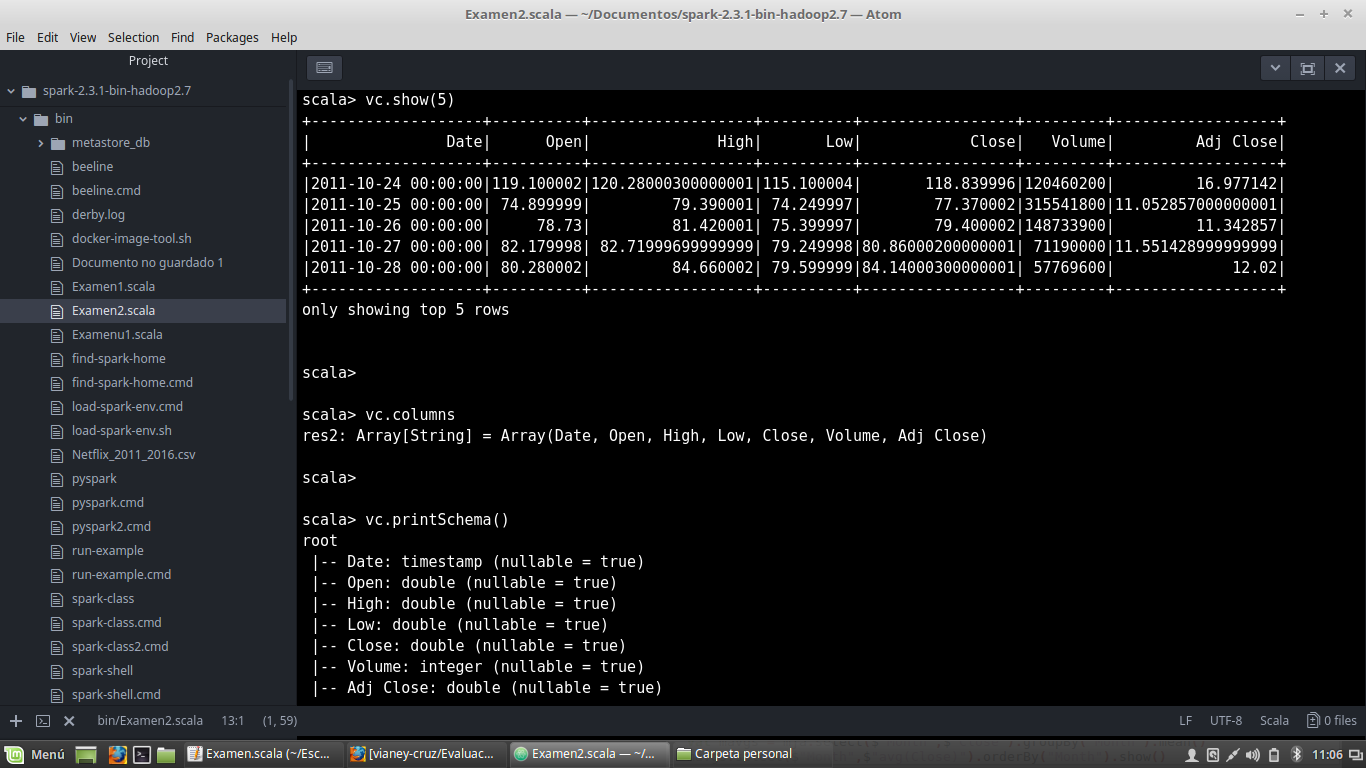
## Cargue el archivo Netflix Stock CSV, haga que Spark infiera los tipos de datos.



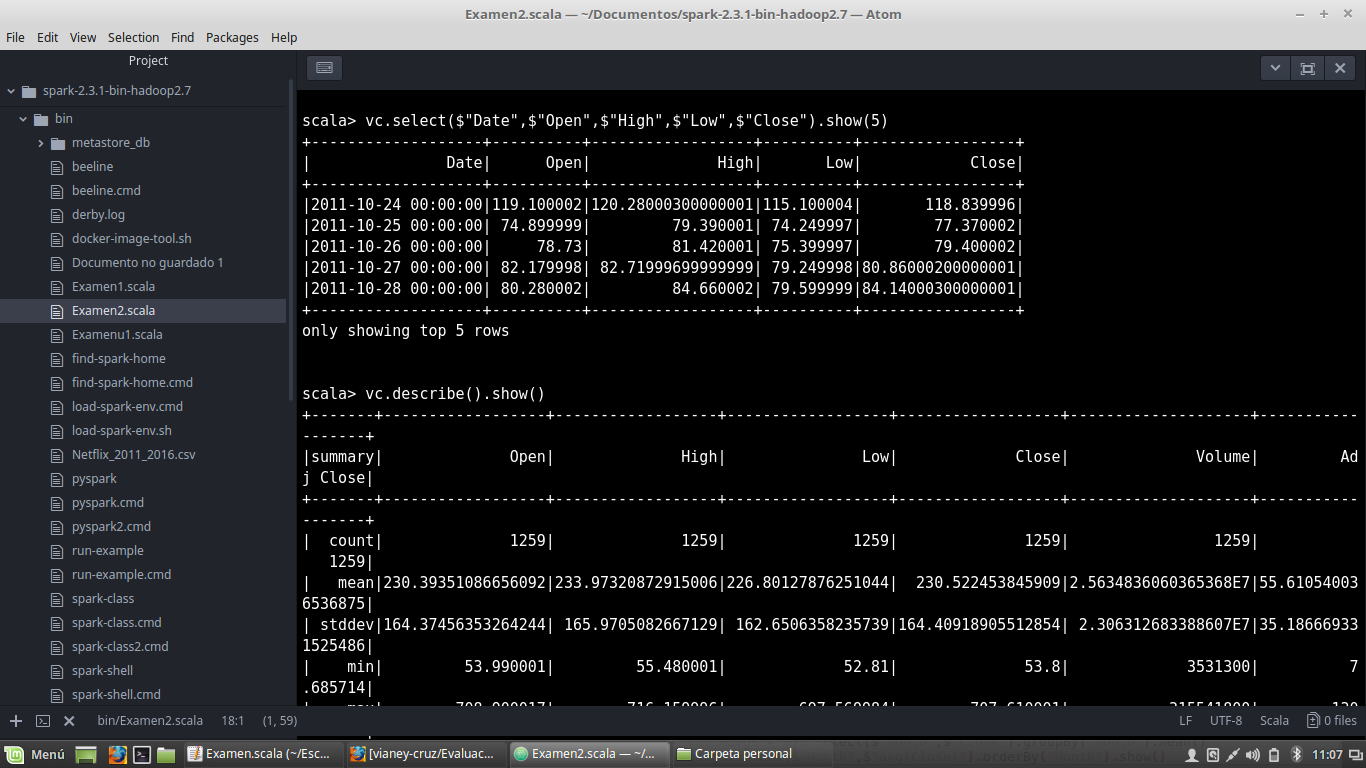
## ¿Cuáles son los nombres de las columnas?



## ¿Cómo es el esquema?



## Imprime las primeras 5 Columnas.

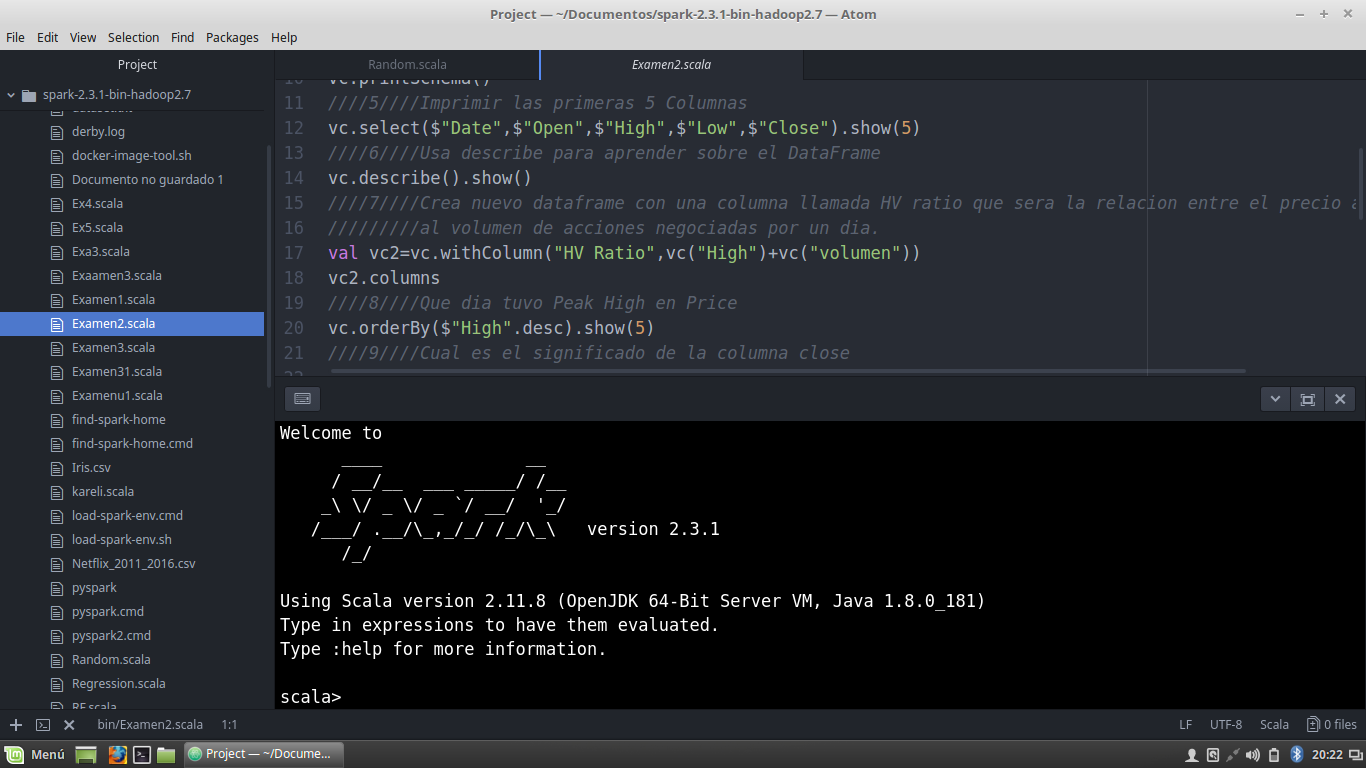


## Usa describe() para aprender sobre el DataFrame.

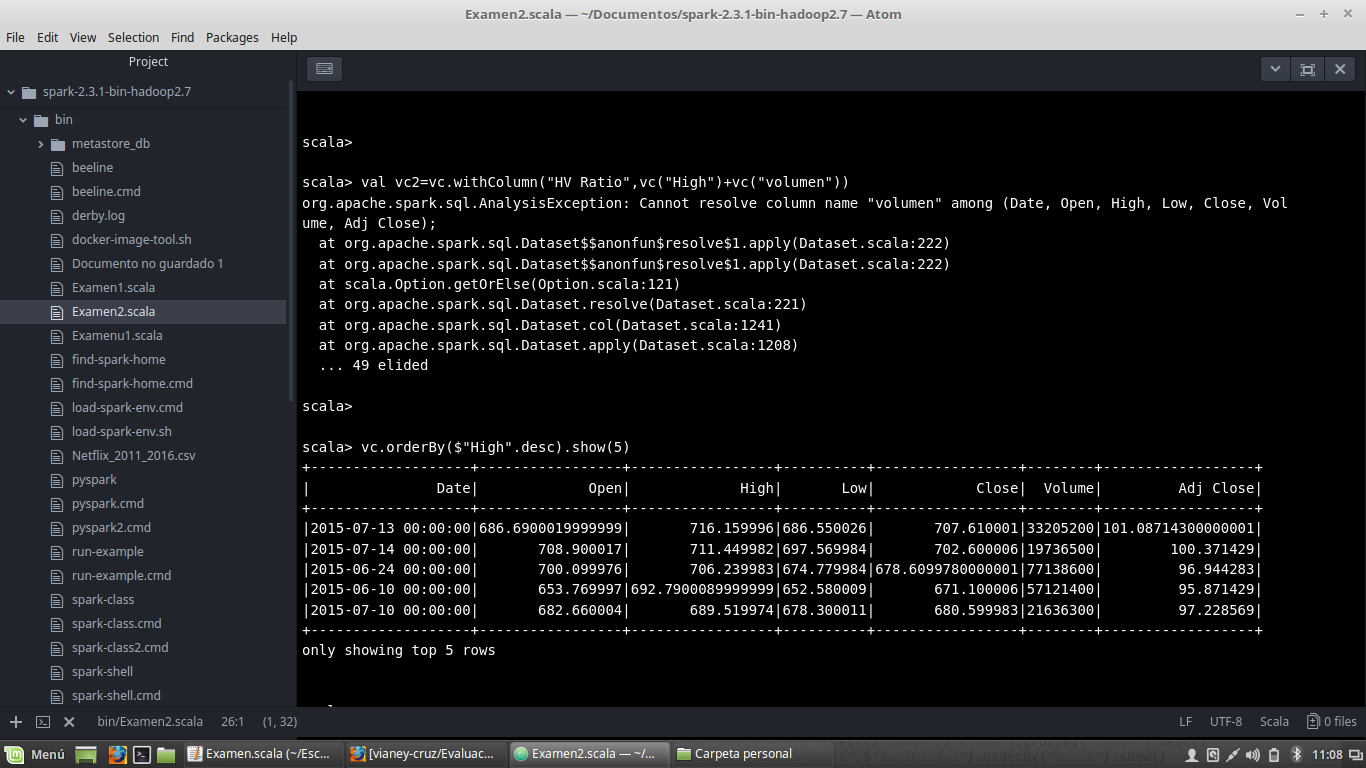
## 

## G:\fotos examen2\4.png

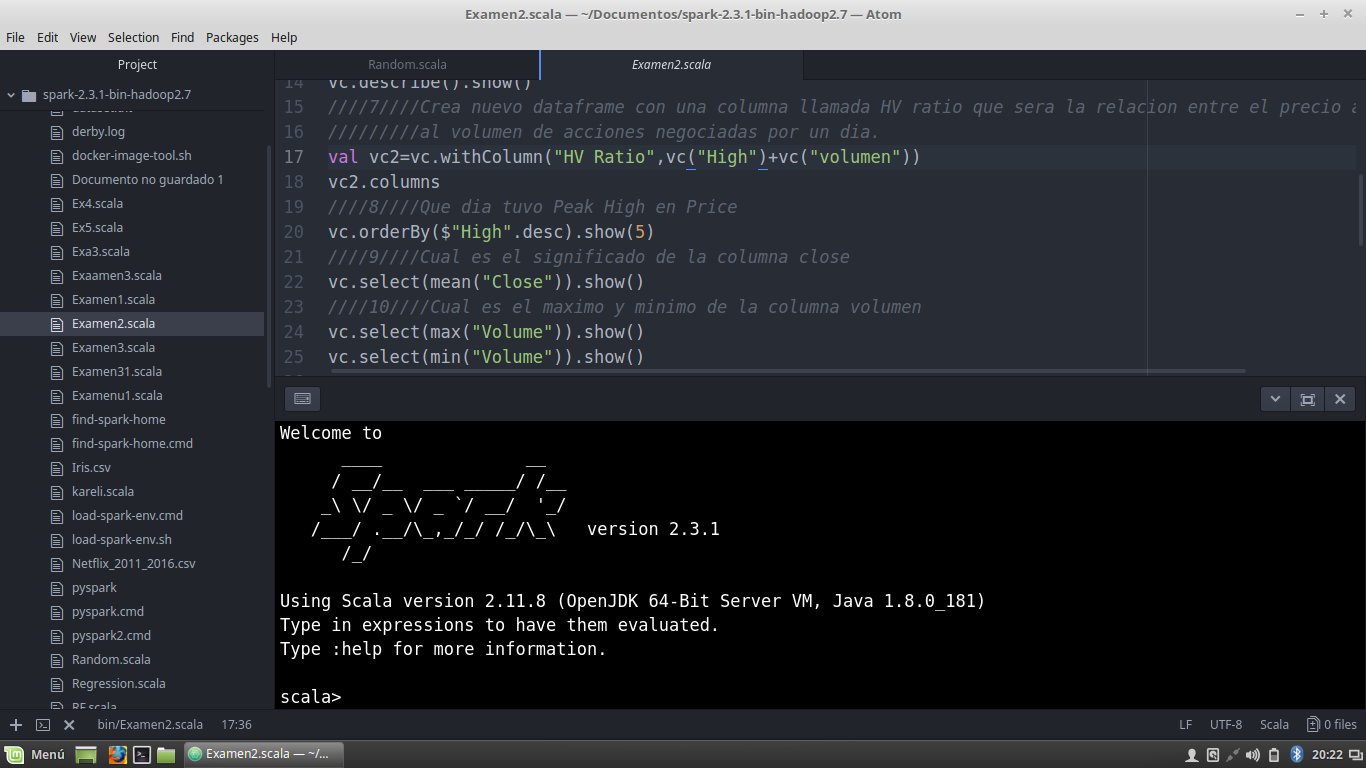
## Crea un nuevo DataFrame con una columna llamada HV Ratio que es la relación entre el precio alto frente al volumen de acciones negociadas por un día.



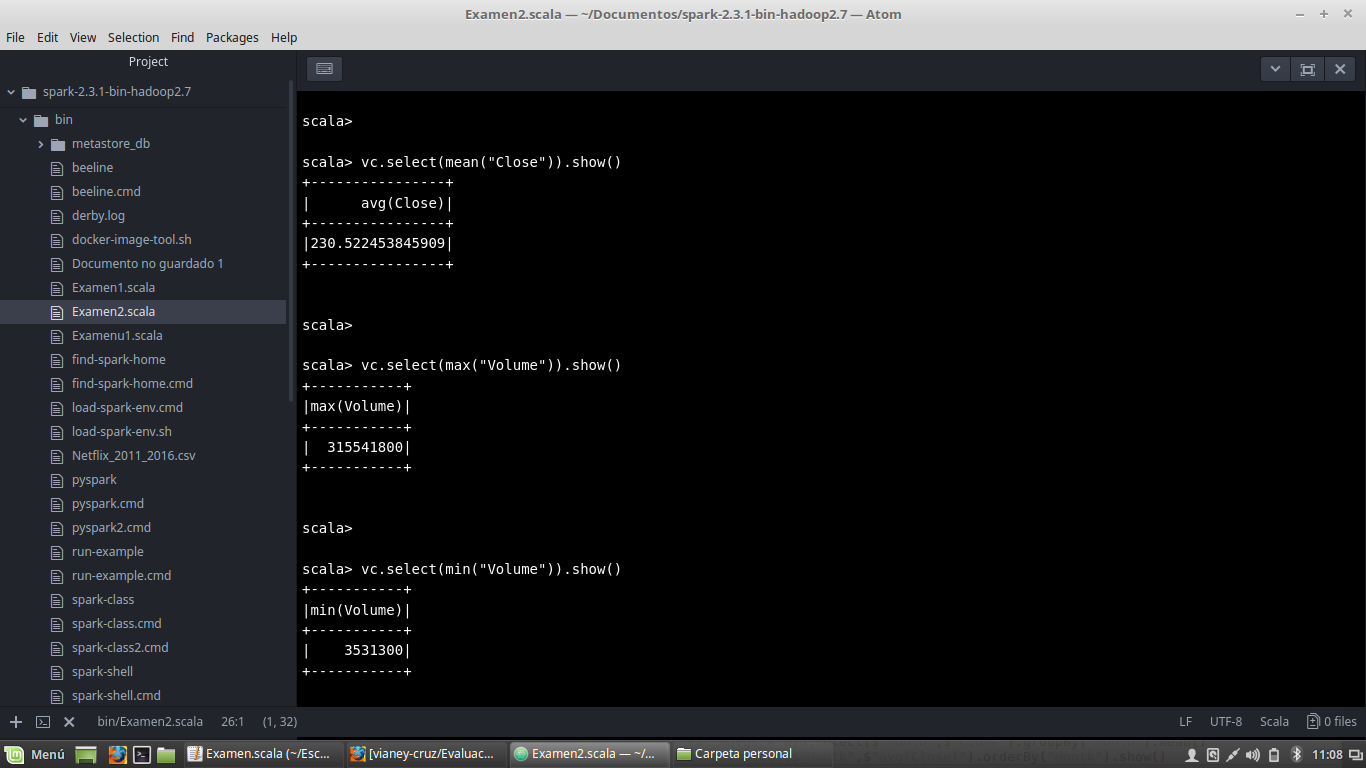
## Que dia tuvo Peak High en Price?



## ¿Cual es el significado de la columna cerrar(close)?

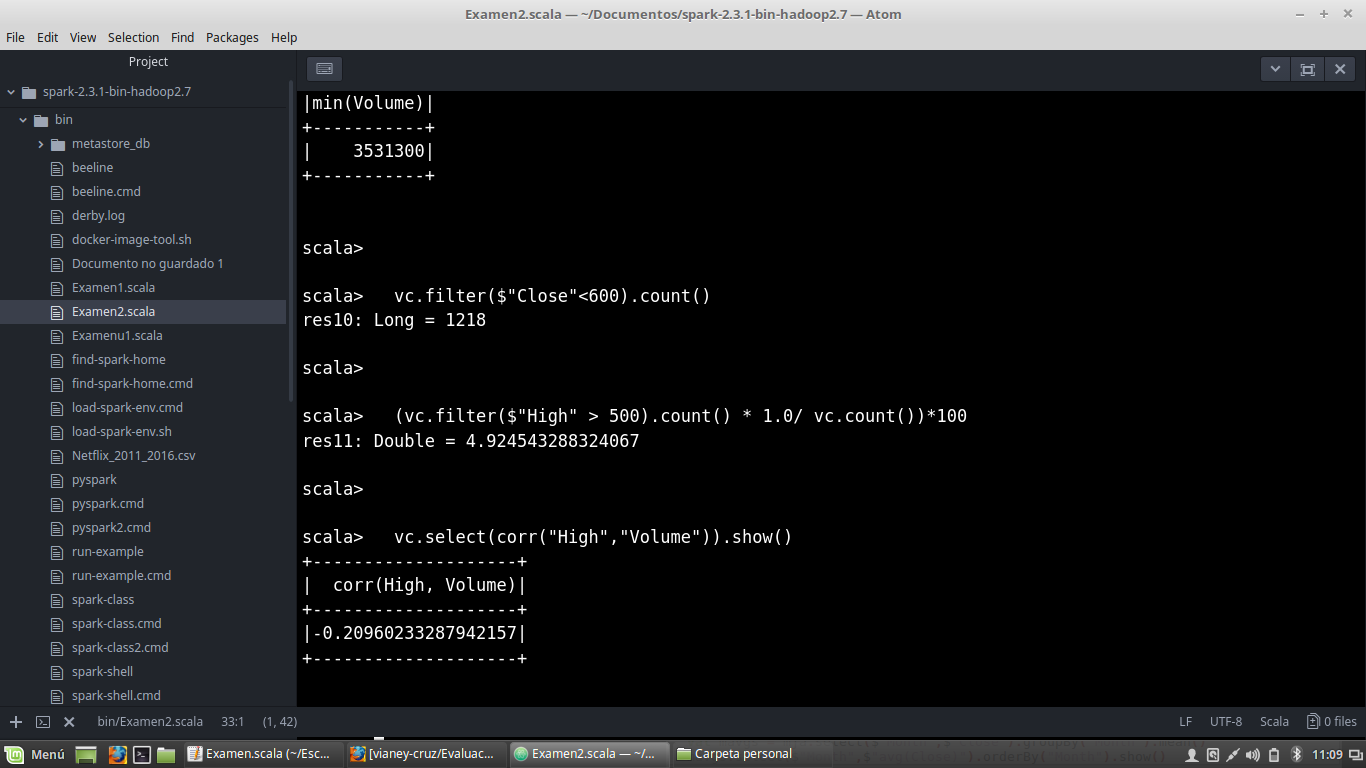


## ¿Cuál es el máximo y mínimo de la columna volumen?

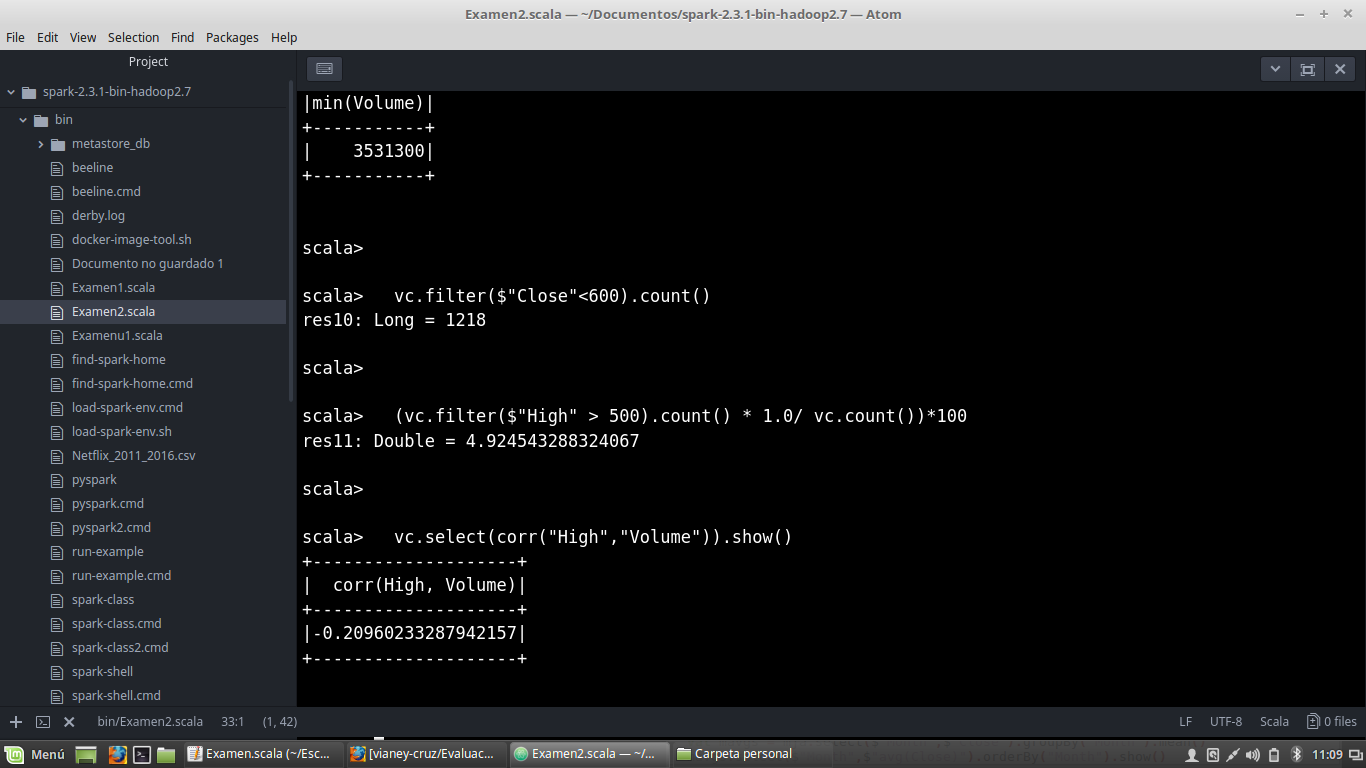


## Con sintaxis Scala / Spark $ conteste lo siguiente:

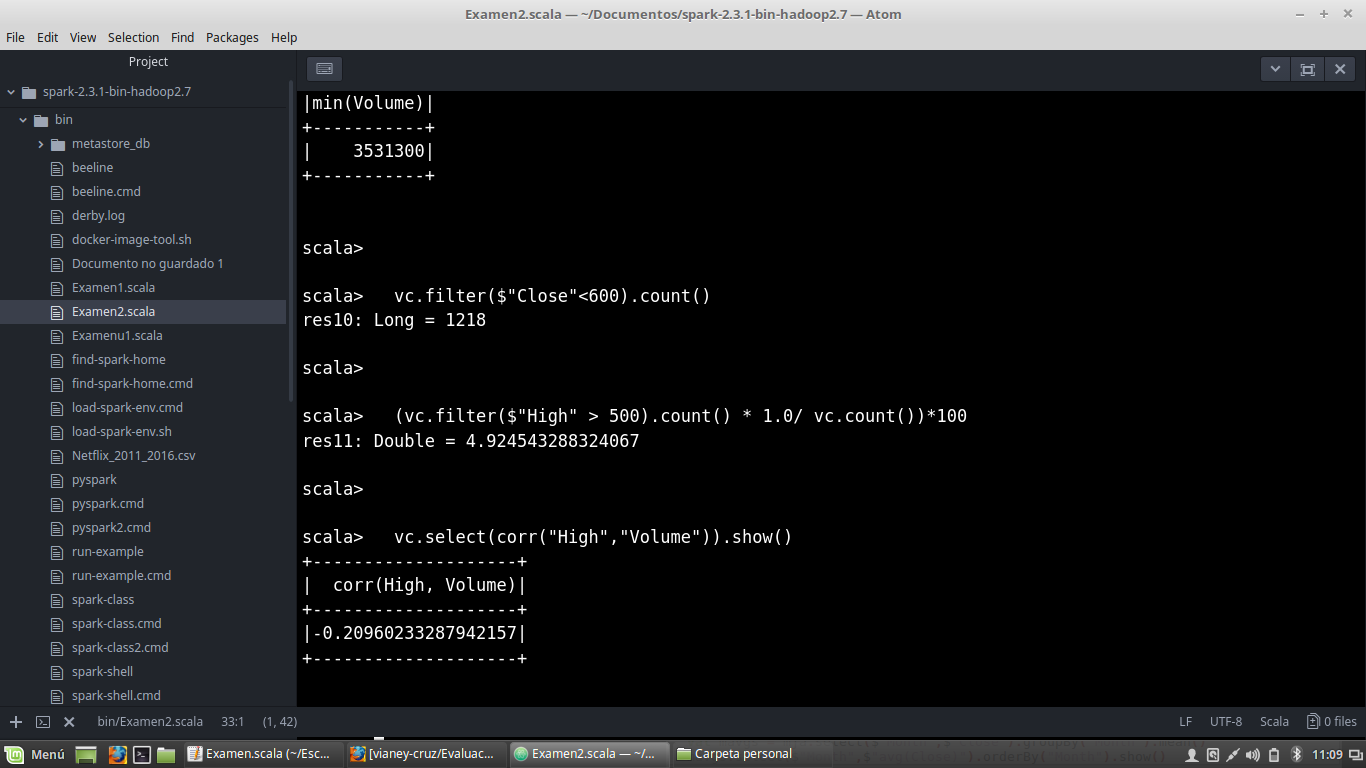
1. ¿Cuantos días fue el cierre inferior a $ 600?



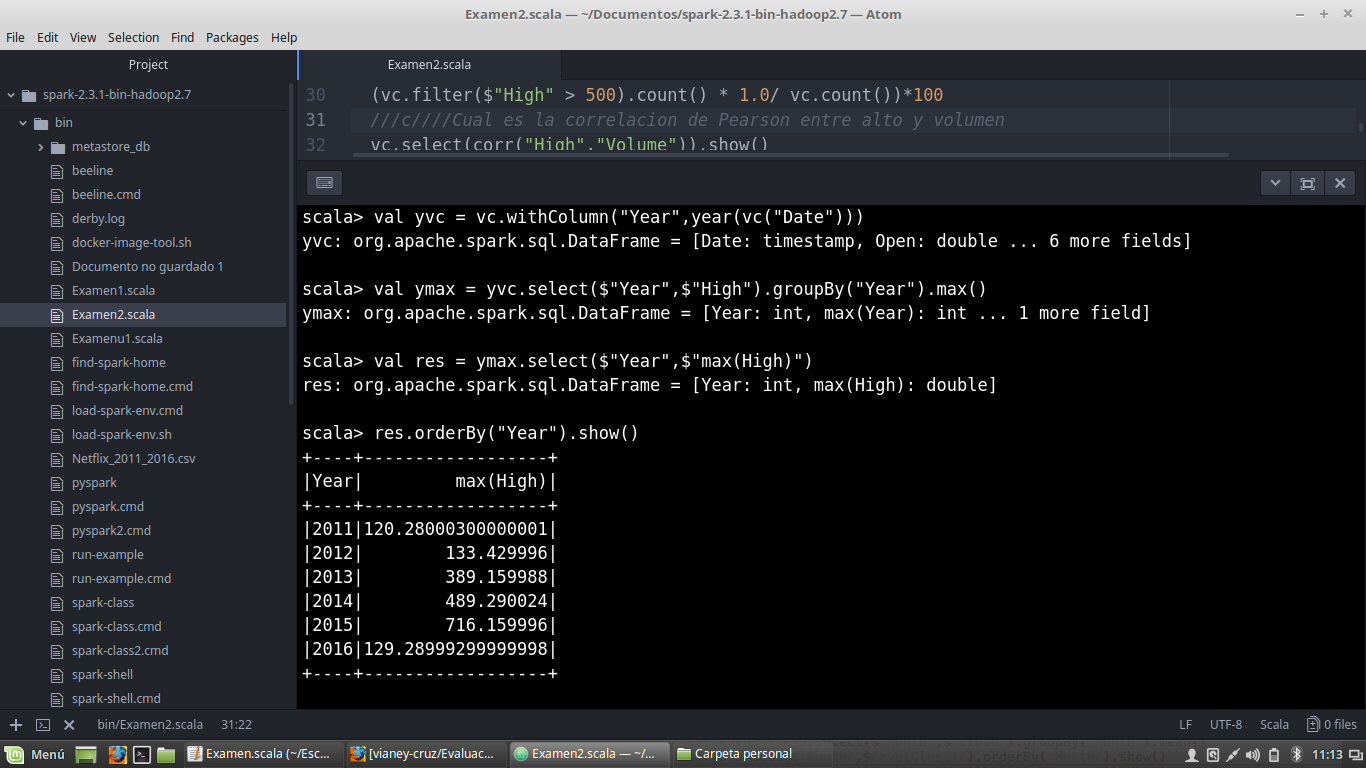
1. ¿Qué porcentaje del tiempo fue el alto mayor de 500?



1. ¿Cuál es la correlación de Pearson entre alto y volumen?



1. Cuál es el máximo alto por a;o?



1. ¿Cual es el promedio de cierre para cada mes del calendario?

