**TALLER 2 ANÁLISIS MULTIVARIADO**

**MAESTRÍA DE SISTEMAS**

**MAYO 4 DE 2019**

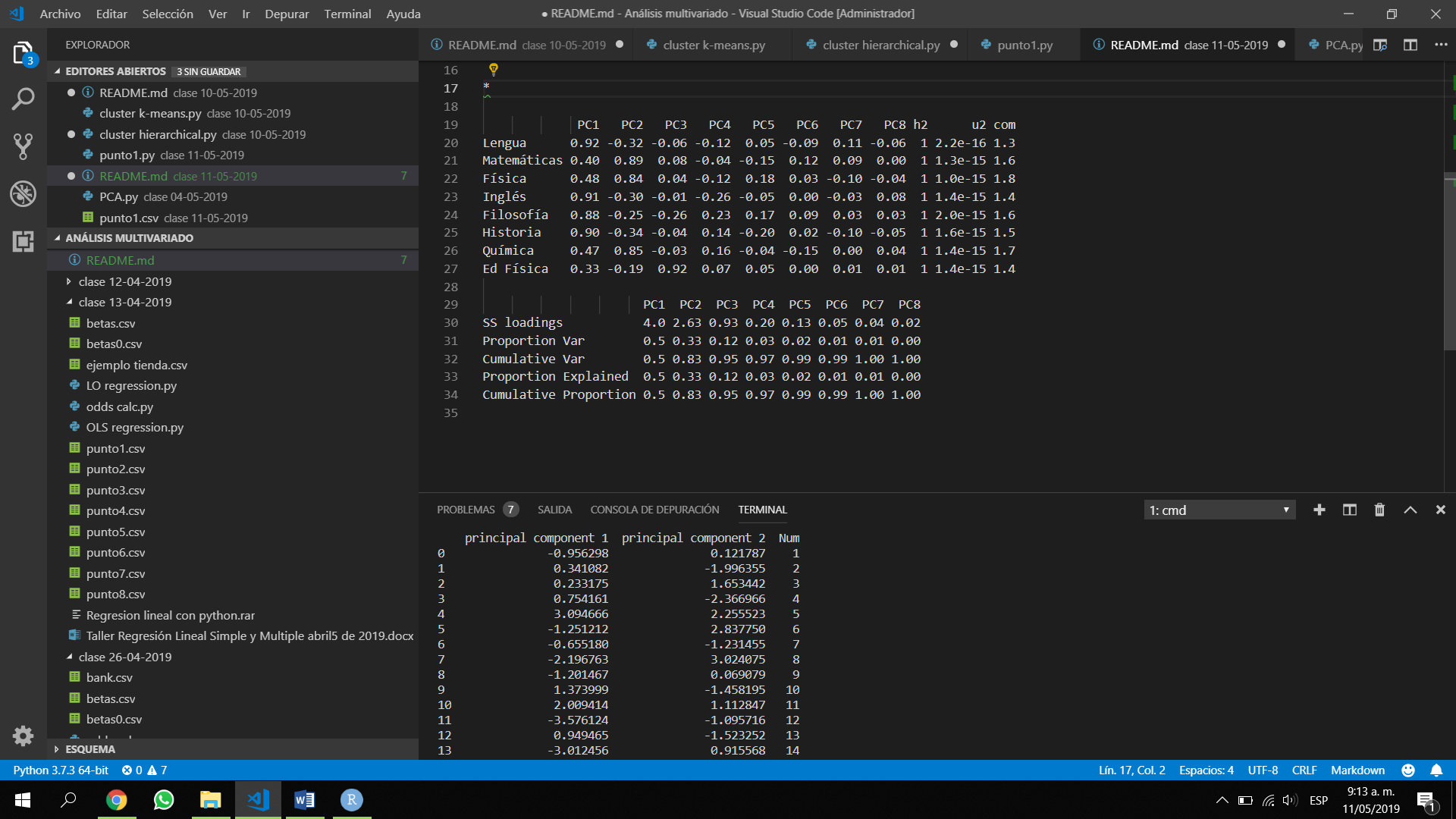
1. Se dispone de los datos de 20 alumnos de un grupo de clases y se tienen las siguientes calificaciones en las asignaturas nombradas.

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | Lengua | Matemáticas | Física | Inglés | Filosofía | Historia | Química | Ed Física |
| 1 | 5 | 5 | 5 | 5 | 5 | 5 | 5 | 5 |
| 2 | 7 | 4 | 3 | 8 | 4 | 7 | 3 | 8 |
| 3 | 5 | 8 | 7 | 6 | 5 | 6 | 7 | 5 |
| 4 | 7 | 2 | 4 | 8 | 7 | 7 | 3 | 6 |
| 5 | 8 | 9 | 10 | 8 | 8 | 7 | 9 | 4 |
| 6 | 4 | 9 | 8 | 4 | 3 | 4 | 7 | 5 |
| 7 | 6 | 4 | 4 | 6 | 5 | 5 | 3 | 7 |
| 8 | 4 | 7 | 8 | 3 | 3 | 2 | 8 | 3 |
| 9 | 5 | 5 | 4 | 5 | 6 | 5 | 5 | 1 |
| 10 | 7 | 4 | 5 | 7 | 8 | 8 | 4 | 6 |
| 11 | 7 | 8 | 8 | 7 | 7 | 6 | 7 | 9 |
| 12 | 4 | 3 | 3 | 4 | 3 | 2 | 1 | 4 |
| 13 | 7 | 4 | 4 | 7 | 8 | 7 | 4 | 5 |
| 14 | 3 | 5 | 5 | 2 | 3 | 3 | 5 | 7 |
| 15 | 5 | 6 | 6 | 5 | 5 | 5 | 6 | 6 |
| 16 | 8 | 9 | 9 | 9 | 9 | 9 | 9 | 8 |
| 17 | 8 | 5 | 5 | 8 | 9 | 9 | 5 | 4 |
| 18 | 7 | 5 | 4 | 7 | 7 | 8 | 5 | 7 |
| 19 | 5 | 2 | 2 | 4 | 5 | 5 | 3 | 6 |
| 20 | 5 | 5 | 3 | 4 | 5 | 5 | 5 | 5 |

Se pide:

1. Realizar un análisis de componentes principales que resuma la información aportada por las variables

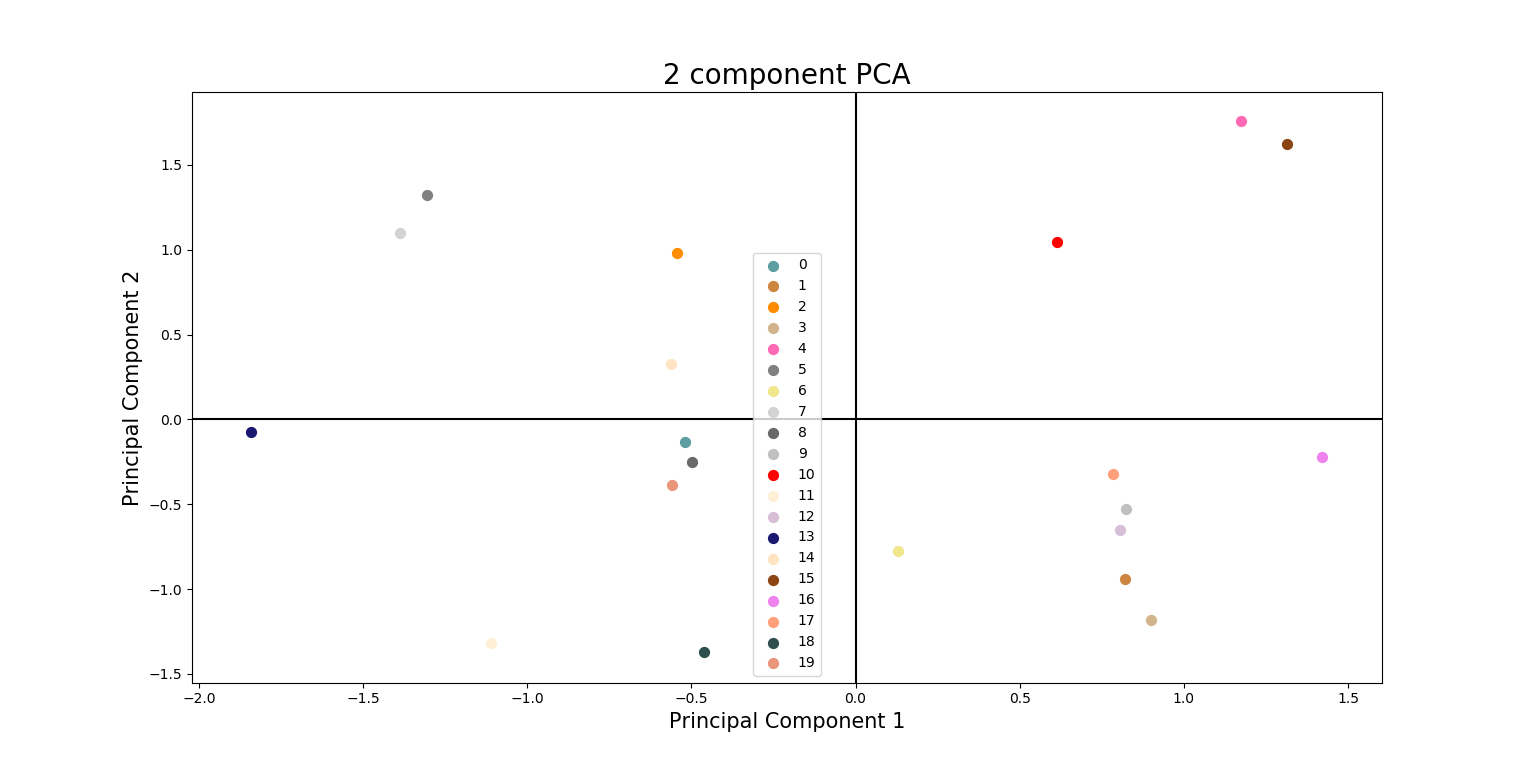
**Análisis de componentes principales sin rotar con 8 factores**



Como se logra ver en el resultado el *Proportion Explained* va decayendo después del factor *PC3*, la explicación por cada materia es más alta en su mayoría para los primeros 2 factores (PC1 PC2) por lo cual será solo con 2 factores, si vemos, así como está los primeros 2 factores explica lo siguiente:

* PC1 explica mejor las materias
  + Lengua
  + Inglés
  + Filosofía
  + Historia
  + Ed Fisica
* PC2 explica mejor las materias
  + Matemáticas
  + Física
  + Química

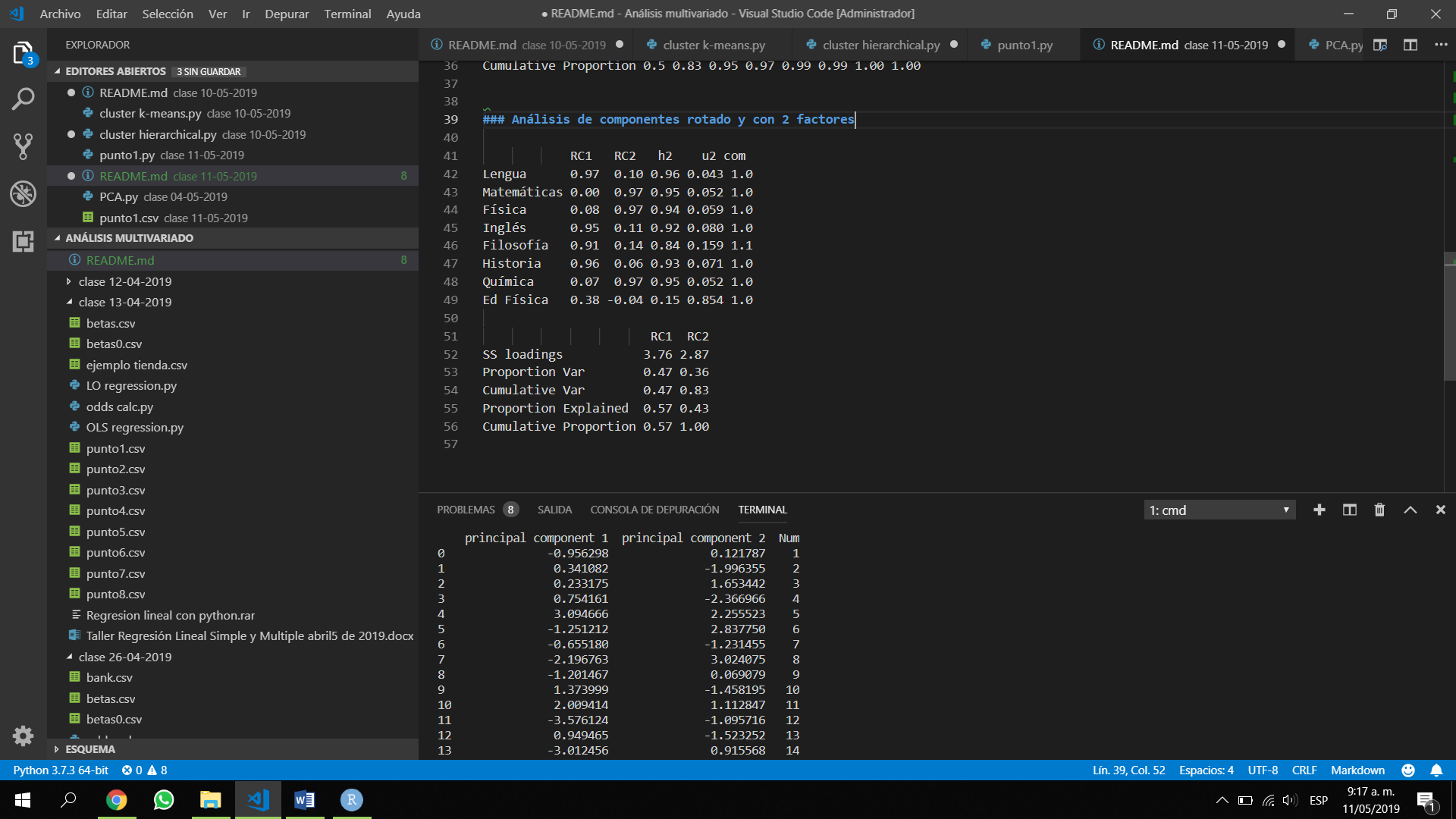
1. Realizar un gráfico de dispersión de los casos una vez encontrado el número de componentes



1. Si se realiza una rotación ¿se acomodan mejor las variables?

Entre más positivo sea es mejor el estudiante en el área relacionada, entre más negativo es peor o “mas malo”

**Análisis de componentes rotado y con 2 factores**



En este caso el *Proportion Explained* es bastante alto (la suma entre estos es 1) por lo cual explica estos 2 factores indican:

* PC1 explica mejor las materias
  + Lengua
  + Inglés
  + Filosofía
  + Historia
  + Ed Fisica
* PC2 explica mejor las materias
  + Matemáticas
  + Física
  + Química

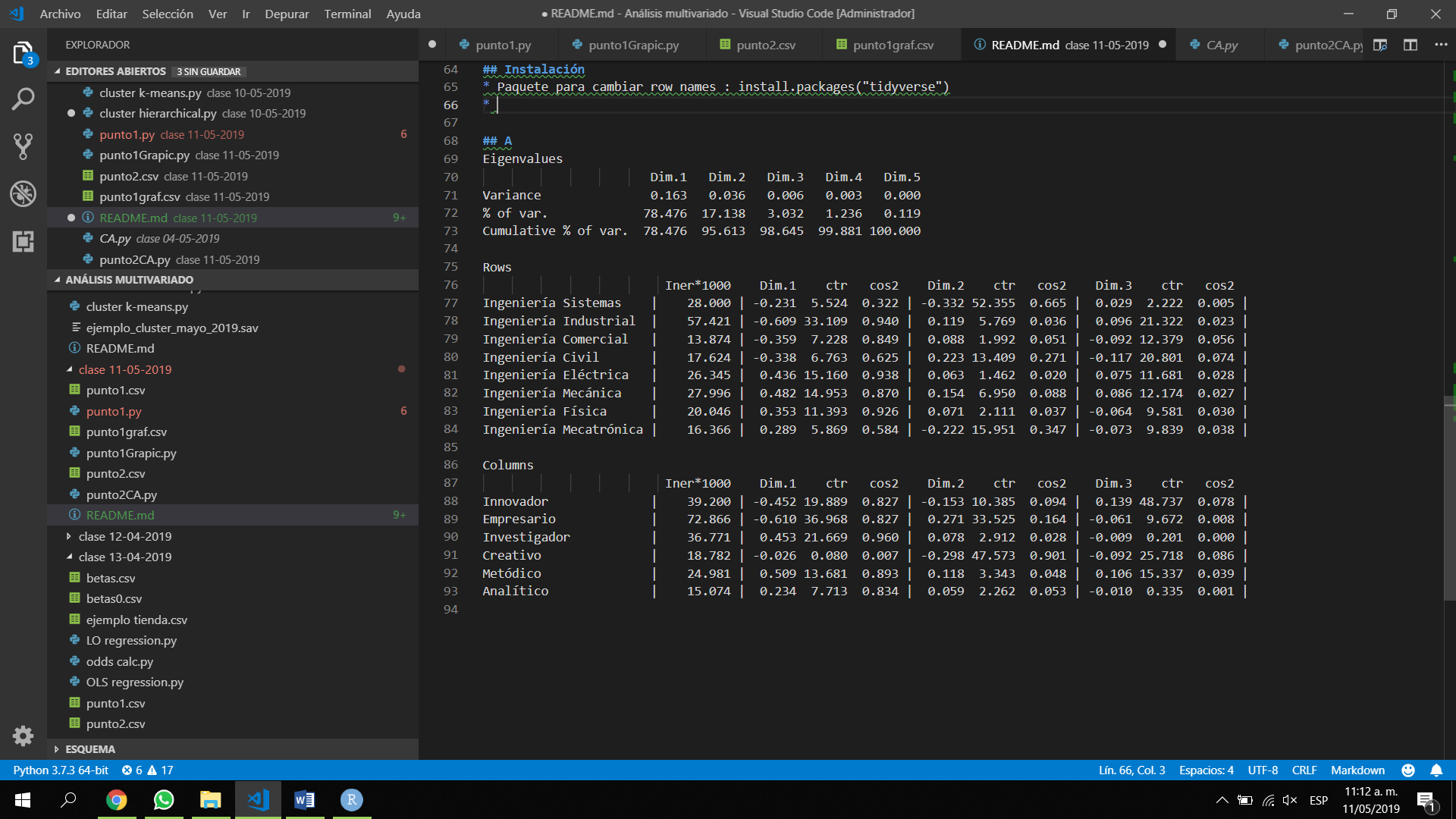
Con la rotación puede explicar mejor con los primeros 2 factores (PC1, PC2)

1. A continuación, se muestran los datos de la apreciación que tienen diferentes gerentes de recursos humanos con respecto a los diferentes perfiles de egresados de las ingenierías que se pueden encontrar en el eje cafetero.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **CARRERA** | **PERFIL O CARACTERÍSTICA DE LA CARRERA** | | | | | |
| Innovador | Empresario | Investigador | Creativo | Metódico | Analítico | |
| Ingeniería Sistemas | 25 | 14 | 11 | 29 | 5 | 16 | |
| Ingeniería Industrial | 24 | 28 | 5 | 12 | 3 | 14 | |
| Ingeniería Comercial | 10 | 15 | 5 | 10 | 2 | 12 | |
| Ingeniería Civil | 9 | 18 | 8 | 9 | 3 | 10 | |
| Ingeniería Eléctrica | 7 | 5 | 20 | 12 | 12 | 21 | |
| Ingeniería Mecánica | 5 | 4 | 17 | 7 | 9 | 20 | |
| Ingeniería Física | 6 | 9 | 22 | 16 | 11 | 24 | |
| Ingeniería Mecatrónica | 8 | 3 | 14 | 18 | 6 | 19 | |

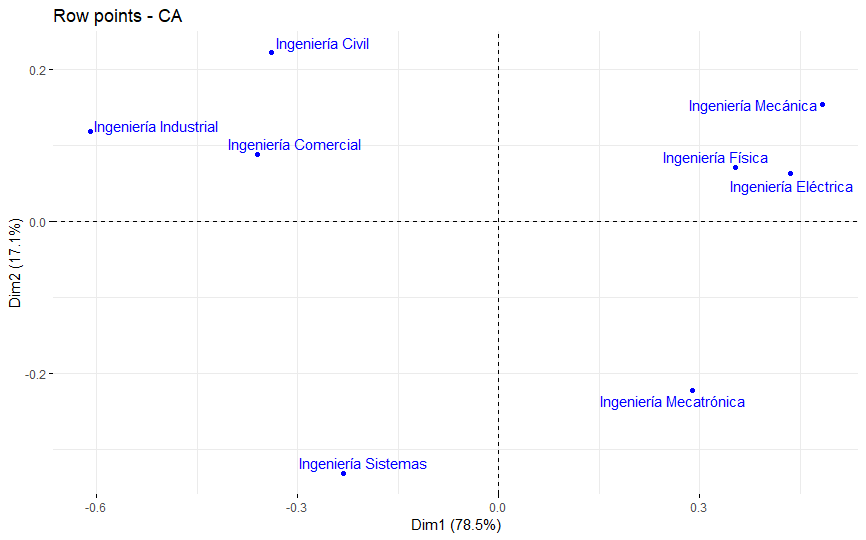
Se pide realizar:

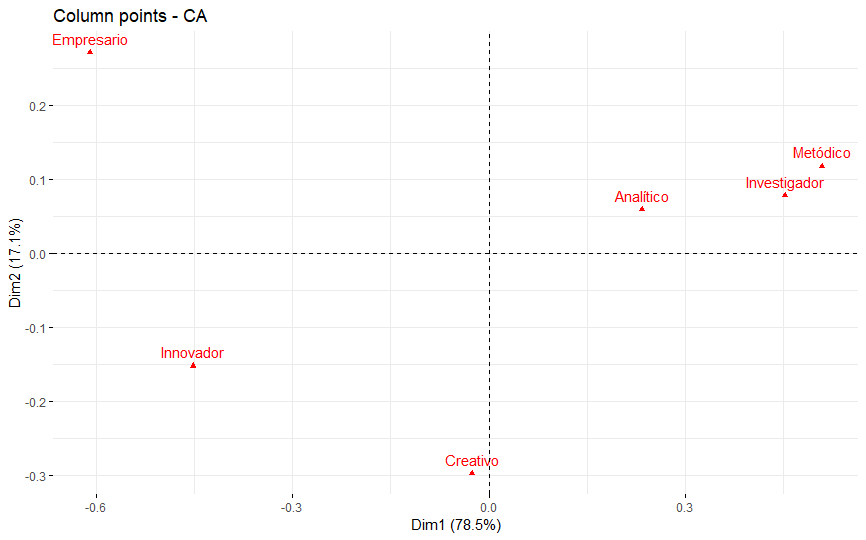
1. Un análisis de correspondencias adecuado para explicar la relación entre características y carreras.



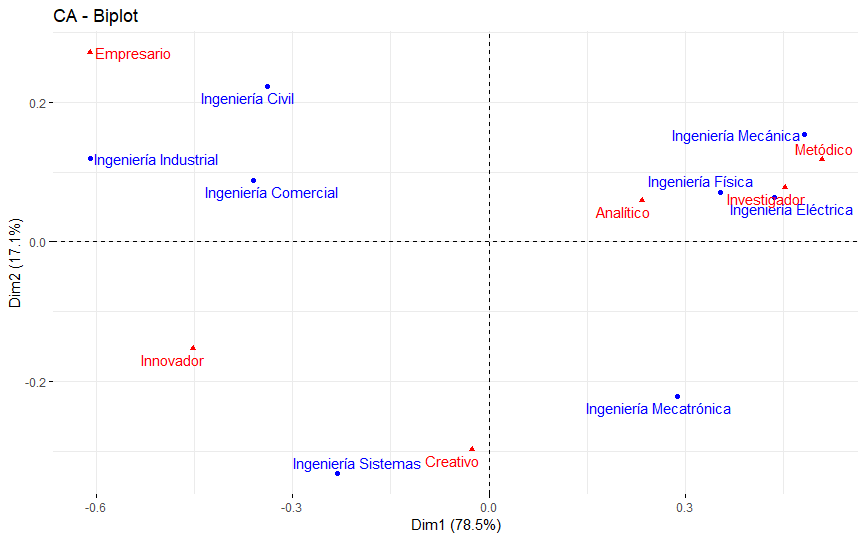
De acuerdo al análisis CA, se logra determinar que ,

1. Un gráfico de dispersión en dos dimensiones tanto para perfil como para carrera





1. Un gráfico de dispersión en dos dimensiones en donde se puede observar la relación entre perfil y carrera



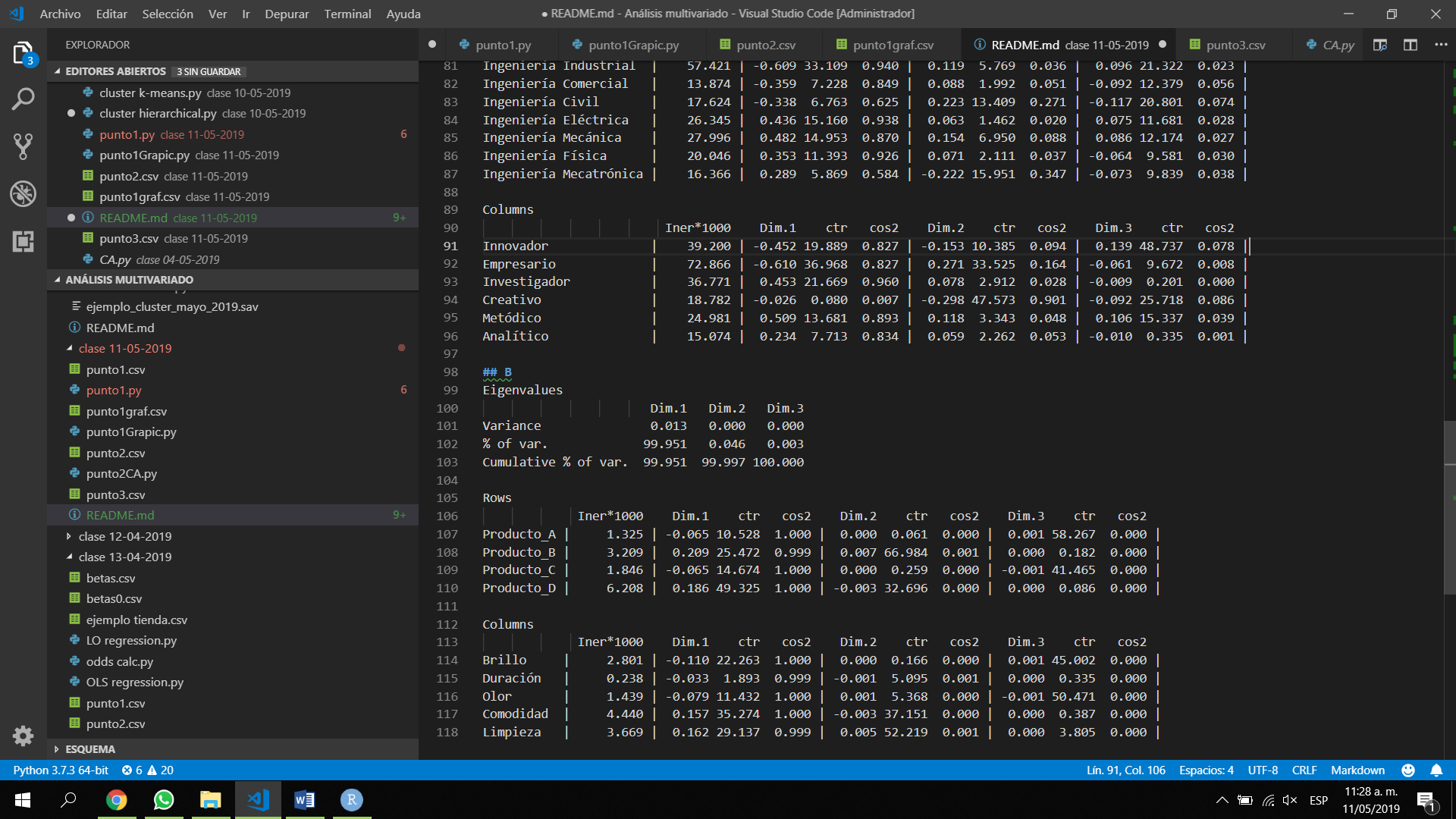
Con el análisis de filas y columnas se logra identificar que los de las diferentes carreras son más que todo:

* Ingenieros de sistemas
  + Creativo
  + Innovador
* Ingeniero Civil
  + Empresario
* Ingenieros industriales
  + Empresario
* Ingeniero Comercial
  + Empresario
* Ingeniero Mecanico
  + Analitico
  + Investigador
  + Metodico
* Ingeniero Fisico
  + Analitico
  + Investigador
  + Metodico
* Ingeniero Electrico
  + Analitico
  + Investigador
  + Metodico

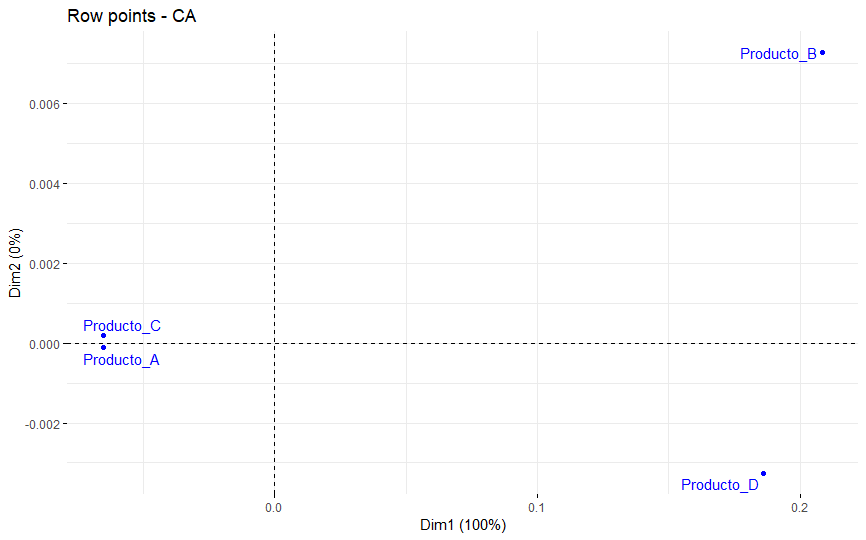
1. Un fabricante de producto de limpieza desea conocer las características más importantes de cada uno de sus productos, para ello realiza una encuesta entre sus clientes. La tabla adjunta muestra las frecuencias de asociación de cada producto con las características analizadas.

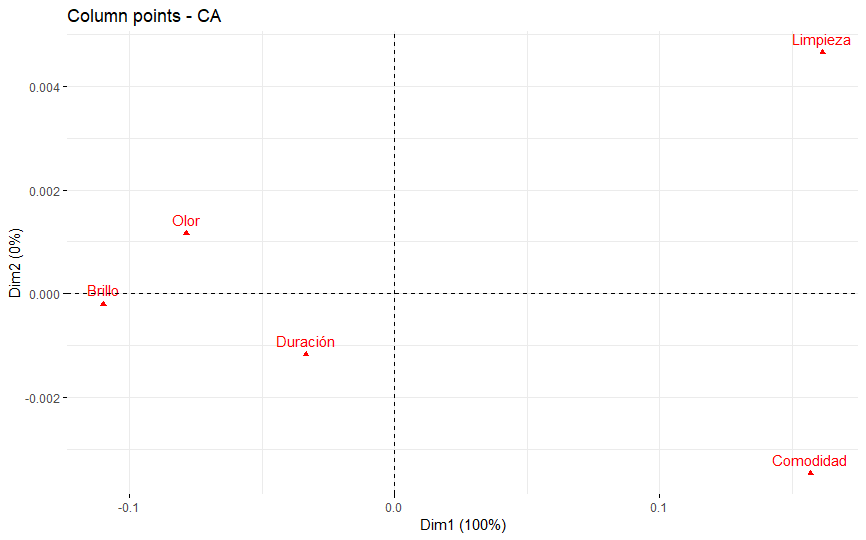
|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  |  | Característica | | | | |  |
|  |  | Brillo | Duración | Olor | Comodidad | Limpieza | Total |
| Producto | Producto A | 68 | 60 | 67 | 45 | 35 | 275 |
| Producto B | 12 | 13 | 13 | 15 | 12 | 65 |
| Producto C | 95 | 84 | 94 | 63 | 49 | 385 |
| Producto D | 30 | 32 | 32 | 36 | 28 | 158 |
|  | Total | 205 | 189 | 206 | 159 | 124 | 883 |

1. Un análisis de correspondencias adecuado para explicar la relación entre producto y característica.

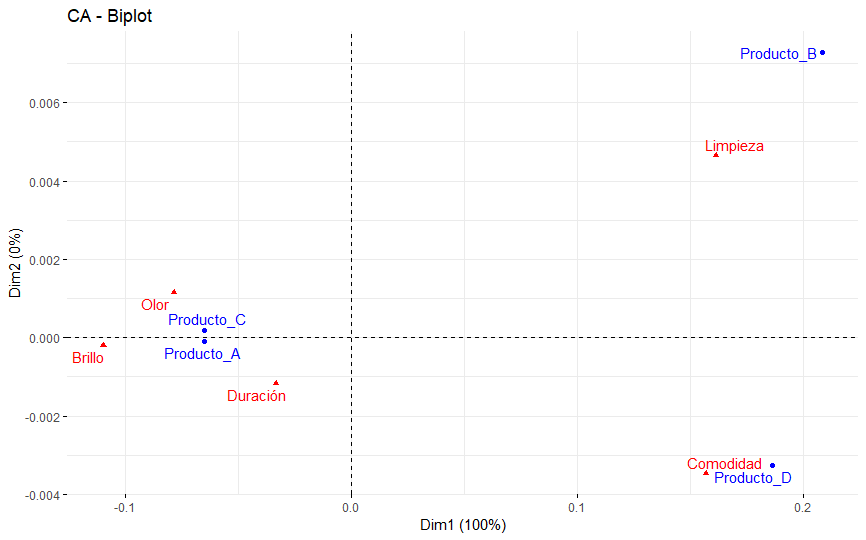


1. Un gráfico de dispersión en dos dimensiones tanto para producto como para característica





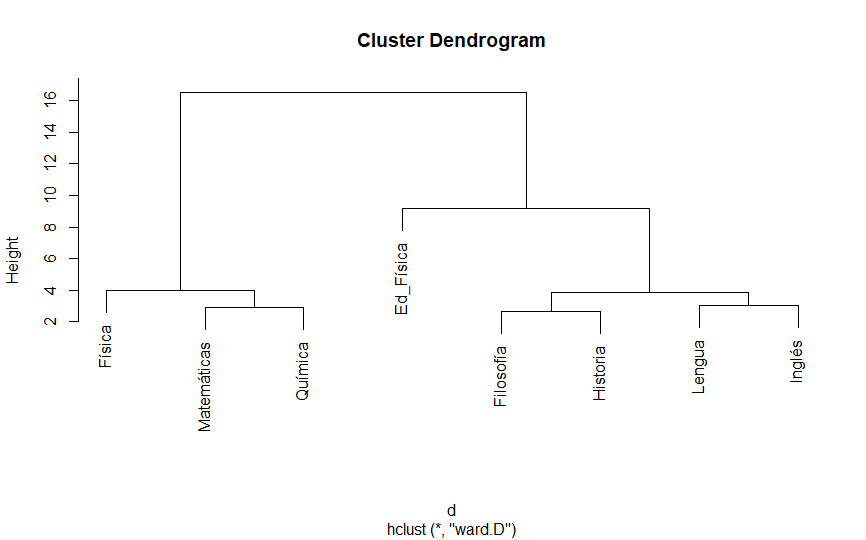
1. Un gráfico de dispersión en dos dimensiones en donde se puede observar la relación entre producto y característica



Cuando se combina en el análisis las filas y las columnas se determina que las características más importantes de los productos son:

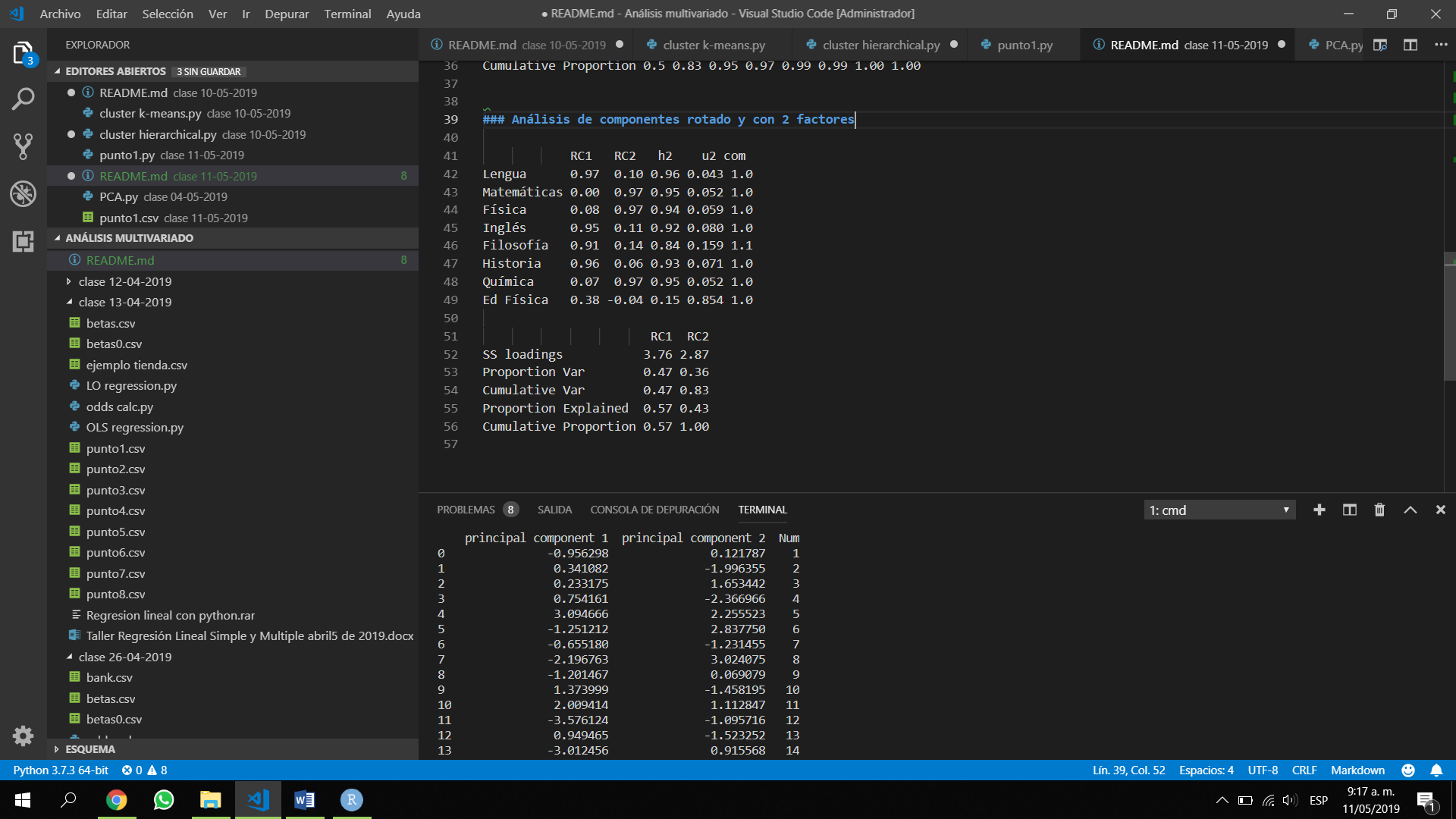
* Producto A
  + Brillo
  + Duración
  + Olor
* Producto B
  + Limpieza
* Producto C
  + Brillo
  + Duración
  + Olor
* Producto D
  + Comodidad

1. Usando el análisis clúster compare los resultados obtenidos del ejercicio 1 donde se aplicó componentes principales y haga un comparativo del dendograma con la matriz de componentes



En el análisis con PCA se determinó que

**Análisis de componentes rotado y con 2 factores**



* PC1 explica mejor las materias
  + Lengua
  + Inglés
  + Filosofía
  + Historia
  + Ed Fisica
* PC2 explica mejor las materias
  + Matemáticas
  + Física
  + Química

De acuero al dendograma para la materia Educación física es la que menos puede explicar el análisis PCA por lo cual tiene mayor Heigth, además en el dendograma también se divide en 2 de la misma forma que lo dividen los factores, es decir:

* Parte 1 del cluster
  + Matemáticas
  + Física
  + Química
* Parte 2 del cluster
  + Lengua
  + Inglés
  + Filosofía
  + Historia
  + Ed Fisica