



Administrivia

75.06 / 95.58 Organización de Datos

Administrivia

- Equipo
- Asistencia
- Correlativas
- Regimen de Cursada
- Trabajo Práctico
- Bibliografía
- Piazza
- Gradiance
- Kaggle

Equipo

- JTP: Natalia Golmar



Equipo

- JTP: Natalia Golmar
- Damián Martinelli



Equipo

- JTP: Natalia Golmar
- Damián Martinelli
- Martín Ramos Mejía



Equipo

- JTP: Natalia Golmar
- Damián Martinelli
- Martín Ramos Mejía
- Joaquín Torré Zaffaroni



Equipo

- JTP: Natalia Golmar
- Damián Martinelli
- Martín Ramos Mejía
- Juan Andrés Laura
- Joaquín Torré Zaffaroni
- Luis (Maurice) Argerich



Asistencia



Alumnos Invitados

- Son bienvenidos
- Si están interesados en obtener un certificado los requisitos son:
 - 80% de asistencia a clases
 - Completar y Aprobar los TPs de la materia
- Se puede inscribir al finalizar esta clase
- No hay examen

Correlativas

- Algoritmos y Programación II es necesaria, fundamental e imprescindible.
- Todas las demás: I don't care.

Conocimientos Previos

- Manejo de Archivos
- Noción de complejidad (orden, etc)
- Hashing
- Árboles binarios, B, B+
- Álgebra II es deseable:
 - SVD (Descomposición por valores singulares)
 - Autovalores y Autovectores
 - Descomposición Espectral
 - Factorización de Matrices

Régimen de Cursada

- Parcial + TP + Final
- Dos recuperatorios para el parcial
- Los que aprueban el parcial en primera o segunda oportunidad pueden rendir un examen por promoción.





Nota Final

$$N_{60-100} = \frac{1}{3}P + \frac{1}{3}TP + \frac{1}{3}C$$

$$N_{4-10} = 1.5N_{60-100} - 50$$

$$N_{prom} = N + 1$$

Material del Curso

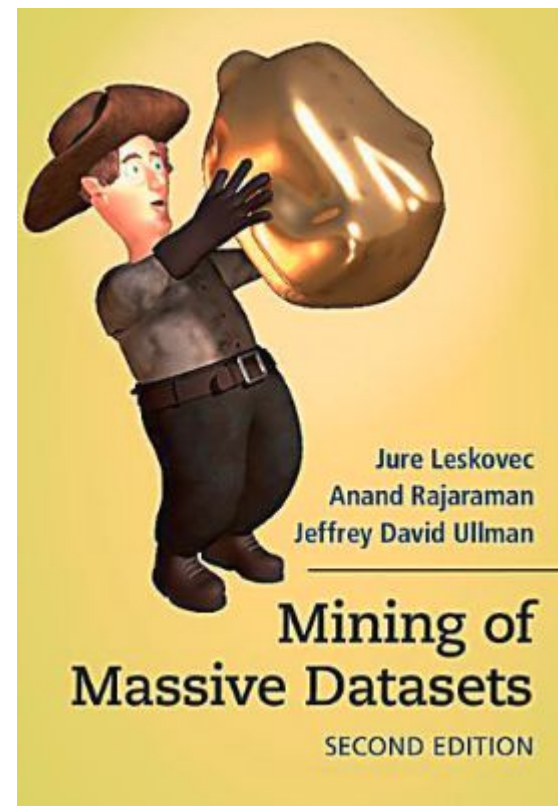
- Apunte (en Piazza)

Contents

I	Data Science	1
1	Fundamentos	3
1.1	Introducción	3
1.1.1	Hacer una pregunta interesante	4
1.1.2	Conseguir los datos necesarios	10
1.1.3	Explorar los datos	14
1.1.4	Aplicar a los datos el algoritmo necesario	18
1.1.5	Comunicar los resultados	18
1.2	Formatos de Datos	18
1.2.1	Data Frames	18
1.2.2	Texto	24
1.2.3	Datos Matriciales	25
1.2.4	Imágenes	26
1.3	Niveles de Almacenamiento	26
1.3.1	Datos en Memoria	27
1.3.2	Datos en Disco	27
1.3.3	Datos en un Cluster	28
1.4	La Ley de Zipf y las Leyes de Potencias	28
1.4.1	Pareto, Zipf y Power-Laws	29
1.4.2	Leyes de Potencias	33
1.4.3	Propiedades de las Leyes de Potencias	34
1.4.4	Origen de las leyes de potencias	36
1.4.5	Resumen	36
1.5	Algunos Elementos de Probabilidad y Estadística	37
1.5.1	El Principio de Bonferroni	37
1.5.2	La Ecuación mas peligrosa de la historia	39
1.5.3	Frecuentistas vs Bayesianos	41
1.5.4	La Única Ley sin Explicación	46
1.5.5	Skewness	47
1.6	Algoritmos Aleatorizados	48
1.6.1	Algoritmo de Fermat	49
1.6.2	Algoritmo de Miller-Rabin	50
1.6.3	Random Walks	50
1.6.4	Markov Chains	51

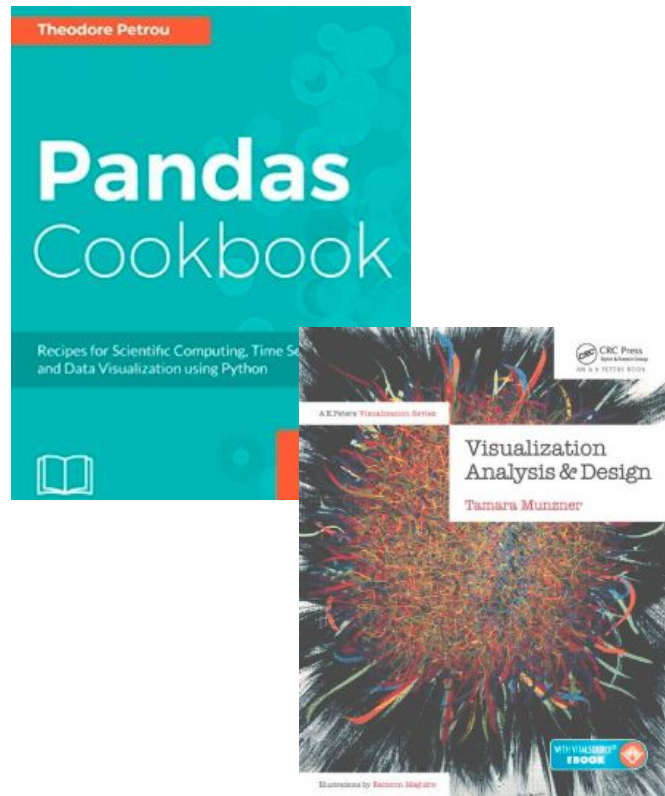
Material del Curso

- Apunte (en Piazza)
- Bibliografía: (todos son gratuitos y online)
 - Mining Massive Datasets [Leskovec,...]
 - Introduction to Information Retrieval [Manning,....]
 - Data Compression Explained [Mahoney]



Material del Curso

- Apunte (en Piazza)
- Bibliografía:
 - Mining Massive Datasets [Leskovec,...]
 - Introduction to Information Retrieval [Manning,...]
 - Data Compression Explained [Mahoney]
- Complementos:
 - Pandas Cookbook [Petrrou]
 - Visualization Analysis & Design [Munzzner]



Material del Curso

- Piazza
- https://piazza.com/universidad_de_buenos_aires/fall2019/7506/home
 - Apunte
 - Calendario
 - Preguntas y Respuestas
 - Enunciado del TP
 - Parciales Anteriores
 - Material de Lectura
 - Y mucho más...

The screenshot shows the Piazza interface for the course "7506: Organizacion de Datos" at the "Universidad de Buenos Aires - Spring 2019". The top navigation bar includes links for "7506", "Q & A", "Resources", "Statistics", and "Manage Class". Below the header, there are buttons for "Syllabus", a pencil icon for editing, and a lock icon. The main content area has tabs for "Course Information", "Staff", and "Resources". The "Description" section contains text about the course's official site. The "Announcements" section has an "Add" button and a yellow box prompting users to add an announcement. The "General Information" section includes a "Facebook" link. The footer contains copyright information for Piazza Technologies, Inc.

7506 Q & A Resources Statistics Manage Class

Universidad de Buenos Aires - Spring 2019
7506: Organizacion de Datos

Syllabus [Pencil Icon] [Lock Icon]

Course Information Staff Resources

Description [Edit]

Este es el sitio oficial correspondiente a la materia 75.06/95.58 de la Facultad de Ingeniería de la Universidad de Buenos Aires correspondiente al primer cuatrimestre de 2019.

Announcements + Add

Add an Announcement
Click the Add button to add an announcement.

General Information [Edit]

Grupo de Facebook
<https://www.facebook.com/groups/457169454398562/>

Copyright © 2017 Piazza Technologies, Inc. All Rights Reserved.

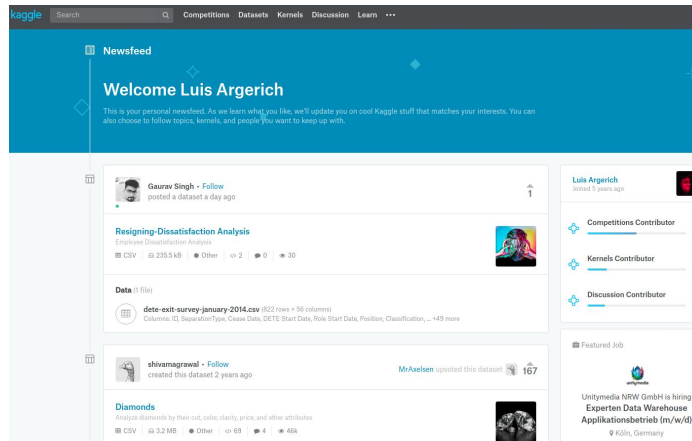
Material del Curso

- Facebook
- 75.06 Organización de Datos
 - Comunicaciones Importantes
 - Links
 - Artículos Interesantes
 - Fingers!
 - Difusión en General



Material del Curso

- Kaggle
- www.kaggle.com
- Es la plataforma para el TP2 del curso (también es útil para el TP1)
- Tiene toneladas de información útil para la materia
- Un usuario por alumno (por favorrrrr!)



Cosas que NO usamos

- Fotocopias (no existe fotocopia alguna sobre esta materia)
- El Campus
- El sitio viejo de Yahoo
- Pascal

Trabajos Prácticos

- Dos TPs
 - Análisis Exploratorio y Visualización de Datos
 - Competencia de Machine Learning
- Grupos de 3 o 4 alumnos
- Python o R
- Nota del TP: Promedio de TP1 y TP2

Herramientas a Instalar

- Python
 - Pandas, matplotlib, numpy, scikitlearn
- Jupyter Lab
- Apache Spark

Código de Honor

- Cualquier Intento de hacer trampa en este curso implica:
 - Quedar libre automáticamente
 - Sumario Administrativo
- Política de Tolerancia Cero



Posts

¡Bienvenidos! Esta es la página del *Laboratorio de Ciencia de Datos* de la Facultad de Ingeniería de la Universidad de Buenos Aires (UBA).

Este sitio funciona como referencia de las actividades (realizadas, en ejecución y planeadas) del laboratorio así como mayor información para los interesados en participar en – o con – el mismo.

También funciona como blog, donde se publicarán pequeñas producciones que tratan sobre temas relacionados al laboratorio. Los artículos serán más informales, teniendo una orientación más cercana a la difusión de las temáticas del laboratorio y las posibles áreas de investigación, siempre manteniendo un nivel de calidad.

Nov 10, 2017

[Reunión del 8/11](#)

Jun 27, 2017

[Primer reunión del LABDAC](#)

May 14, 2017

[¿Qué es la Ciencia de Datos?](#)

¿Preguntas?