Facultad de Ciencias Exactas, Ingeniería y Agrimensura
Licenciatura en Ciencias de la Computación
Probabilidad y Estadística
Trabajo práctico: Estadística descriptiva
Año 2025

Equipo docente: **Lic. Maite San Martín**
Dr. Gustavo Galizzi
Katherine Sullivan

La UNR, universidad estatal en la que usted se está formando actualmente, realiza un convenio con la Organización de Naciones Unidas (ONU) mediante el cual estudiantes avanzados trabajan en entidades públicas con el fin de devolver a la sociedad lo que ésta invirtió para su formación. Durante el cursado de Probabilidad y Estadística en el 3er año de la LCC usted ingresa como data scientist en el sector de Business Intelligence Conosur de la ONU. El sector está liderado por el gerente Rubén Feffer y está organizado de la siguiente forma:

- Advanced Analytics, son 4 data scientists (un economista, un actuario, un ingeniero en sistemas, los tres con una maestría en ciencia de datos, más un PhD en física) los que poseen entre 3 y 12 años de experiencia en ciencia de datos. Usted y sus dos compañeros de equipo, los más junior de este área, aún no están graduados de la LCC. Todo este sector reporta directamente a Rubén Feffer: advanced analytics es el área con la que él más interactúa;
- Data Engineering, 7 integrantes, todos con sólida formación en sistemas (hay especialistas en Big Data, en cloud computing, en arquitectura de data lakes, en infraestructura y en MLOPs que se encargan de poner en producción los modelos desarrollados por los data scientists);
- Data Verticals, 10 personas provenientes de las más diversas carreras, parte de ellos son data analysts;
- Reporting, 6 personas que mantienen los reportes periódicos y además hacen análisis simples ad-hoc;
- Data Governance, 1 vigilante.

El sector ha existido en la organización por más de 25 años, evolucionando con diversos nombres y estructuras. Decididamente usted no ha llegado como el salvador del sector que viene a reinventar la forma de trabajo y a reescribir todo de cero. El sector ya tiene una metodología de trabajo, decenas de repositorios de datos, cientos de scripts, ecosistema en el cual usted debe aprender a manejarse y agregar valor.

Todo el sector considera muy importante la reutilización de código. Todas las mejoras, previamente documentadas, son bienvenidas. Reescribir de cero un script que ya está en el repositorio y es eficiente es visto como muy antieconómico en el sector.

Luego de tres meses aprendiendo del ecosistema, les encargaron el proyecto de barrios populares en Argentina. Antes de conocer la consigna específica que les asignaron, es necesario conocer a los actores.

Rubén Feffer, de 38 años, es el gerente de business intelligence de la compañía, es su jefe directo y fue quien lo contrató hace tres meses.

Rubén posee un título de Licenciado en Estadística, desde su graduación se dedicó a la asistencia técnica en la elaboración de diagnósticos para intervenciones barriales en la ciudad de Rosario y, posteriormente, en Buenos Aires. En el año 2012 cursó una maestría en ciencia de datos e ingresó a trabajar a la organización en el año 2017; ya en 2020 lo ascendieron a gerente del área Business Intelligence. Rubén es una persona muy metódica y organizada, pausado en su hablar, elige sus palabras con gran precisión, reflexivo, considera muchas opciones antes de tomar una decisión, ante una situación difícil de resolver escribe un cuadro en su excel con las alternativas a las que les estima una probabilidad.

Rubén no transparenta fácilmente sus emociones. En las reuniones va escribiendo la minuta en tiempo real y la disponibiliza a todos ni bien termina. Es una persona muy focalizada y ninguna idea foránea lo distrae del problema que debe resolver. Prefiere aprender en forma estructurada y abordar los temas desde lo abstracto. Rubén posee un elevado sentido de la ética y la justicia. A la hora de gestionar su área, necesita una fundamentación para cada una de las decisiones tomadas.

Desde 2023 Rubén reporta directamente a la presidenta del Consejo Económico y Social, ya que se están trabajando simultáneamente varias líneas de trabajo de diagnóstico para implementar estrategias de desarrollo comunitario. En las reuniones semanales que Rubén mantiene con su casi jefa, Ana Priestly, contándole el avance de los múltiples proyectos del sector, ella le pide información adicional sobre el proyecto de barrios populares de Argentina y Rubén le solicita a usted hacer esa presentación en un video para que empiece a ser conocido en la organización.

Ana Priestly, argentina, 48 años, dos hijas gemelas pre adolescentes, es la presidenta del Consejo Económico y Social desde hace dos años y medio, y en su meteórica carrera se pronostica que llegará a la presidencia general de la ONU en dos años más.

Ambos padres de Ana son nacidos fuera de Argentina dedicados en su momento a la actividad consular. Ana emigró de joven, concluyó sus estudios secundarios en el UWC Atlantic College en Gales, se graduó con honores en Ciencias Políticas en la Sorbonne Université de París y cursó una maestría en economía en la London School of Economics and Political Science.

Ana practica yachting desde su infancia, actividad fomentada por su padre quien le inculcó el trabajo en equipo y la competitividad. En su juventud participó de varias competencias internacionales, siendo un punto de inflexión en su vida la accidentada carrera de 1998 54th Sydney to Hobart Yacht Race. En su oficina posee un cuadro de muy importantes dimensiones con una fotografía de esa carrera en donde se aprecia a una joven Ana formando parte de un numeroso equipo sobre una embarcación; al pie del cuadro reza una enorme leyenda “Las regatas se ganan en tierra”.

Ana se unió desde muy joven a la ONU en Europa, estuvo a cargo de la Comisión sobre Población y Desarrollo y la convencieron de hacerse cargo del Consejo Económico y Social para liderar una transformación radical.

Aunque va con una sonrisa y su tono de voz es muy bajo y sereno, todos tienen una especie de temor hacia ella. Se dice que cuando luego de una exposición Ana le dice al

disertante "buen trabajo" sonriendo, antes de los tres meses esa persona ya no está más en la organización.

Usted jamás ha participado en una reunión con ella y ésta será su gran oportunidad de ser conocidx. Ana busca ideas revolucionarias que le permitan aumentar la posición de liderazgo que ya ostenta la organización. Un total de 2500 personas dependen indirectamente de Ana, estando el grueso en la Comisión de Prevención del Delito y Justicia Penal, la Comisión de Ciencia y Tecnología para el Desarrollo y la Comisión para el Desarrollo Social.

Consigna 1 (grupal): Rubén espera de su equipo un informe técnico de no más de 5 páginas en el que se desarrolle un plan de análisis de datos. Este no es el examen final de la asignatura Probabilidad y Estadística, por lo que ustedes no deben explicar la teoría vista en clases, sino que deben contarle a Rubén cómo planifican realizar el análisis que le va a mostrar a la presidenta del Consejo Económico y Social justificando sus decisiones de manera técnica. Ustedes deben ir al grano con Rubén, pero sorprenderlo.

En este documento debe incluir una breve descripción del problema, de la población objeto de estudio, indicar qué variables van a incluir en su análisis y por qué, e indicar qué recursos gráficos y numéricos van a utilizar para cada variable (como mínimo se requiere una variable de cada tipo)[1]. Debe quedar claro por qué el análisis descriptivo sobre este recorte es fundamental para Miranda y para la ONU, haciendo hincapié en el valor agregado de su estrategia de análisis.

Criterios para la evaluación del informe a Rubén	
Porcentaje	Concepto a evaluar
20%	Audiencia. El informe tiene en cuenta la audiencia para la cual está dirigido y saca provecho de las características únicas de esa audiencia.
20%	Historia. Se define de forma clara y justificada el recorte a analizar, resaltando el valor de la propuesta. Esto incluye tanto al universo bajo análisis (si se analiza el total de las viviendas del país o un recorte territorial o dado por alguna característica) así como las variables bajo estudio (selección de algunas sobre las 118 variables presentadas en la tabla de datos).
50%	Contenido. Se cumple la consigna en cuanto a cantidad y tipo de variables. Las técnicas de análisis propuestas son las apropiadas para cada variable y tipo de análisis. Se incluyen las justificaciones pertinentes.
10%	Originalidad del contenido. Las ideas presentadas son originales, ingeniosas, basadas en una profunda comprensión del problema.

Consigna 2 (individual): Usted debe realizar un video presentación a Ana que no puede exceder los 5 minutos en donde le presente las características de los barrios

[1] Es decir, al menos una variable categórica medida en escala nominal, una variable categórica medida en escala ordinal, una variable categórica de respuesta múltiple, una variable cuantitativa discreta y una variable cuantitativa continua.

populares de Argentina. Conocer las características de estos territorios vulnerables es fundamental para poder emprender políticas públicas que aborden sus problemas.

Ana no sabe (ni le interesa saber) sobre estadística ni ciencia de datos pero sí está absolutamente convencida de las posibilidades que brindan las herramientas de la estadística descriptiva para conocer la realidad de estos territorios.

Debe quedarle muy claro que el video a Ana no es contarle lo mismo que le contó a Rubén. Ana le está pidiendo las características de los barrios, mientras que Rubén le está pidiendo que justifique por qué eligió esas herramientas de la estadística descriptiva para realizar la caracterización.

Criterios para la evaluación del video a Ana	
Porcentaje	Concepto a evaluar
10%	Entretenimiento. El video presentación es apasionante y no hay parte que aburra. La atención del espectador debe mantenerse durante los 5 minutos.
15%	Audiencia. El video presentación tiene totalmente en cuenta la audiencia para la cual está dirigido y saca provecho de las características únicas de esa audiencia.
20%	Historia. La presentación narra una historia, hay una clara introducción con un “gancho” que invita a ver el video, un desarrollo adecuado con una continuidad argumental lógica y un desenlace concreto. La narrativa está organizada en torno a las etapas de la pirámide de Freytag o estructura similar.
40%	Consistencia del contenido. Lo presentado refleja fielmente el conocimiento descubierto en el análisis y las conclusiones están sustentadas en datos que aparecen presentados adecuadamente.
10%	Originalidad del contenido. Las ideas presentadas son originales, ingeniosas, basadas en una profunda comprensión del problema.
5%	Transparencia. Es posible acceder al repositorio del grupo y reproducir los análisis realizados.

Para la consigna 1 cada equipo deberá entregar un informe, es una actividad grupal. Para la consigna 2 cada estudiante debe hacer un video presentación, es una tarea individual.

Por más que se hayan formado grupos de dos o tres personas para la consigna 1, los videos deben ser distintos así como el material que los soporta (diapositivas o similares), aunque los recursos descriptivos (gráficos, tablas, resumen numérico) pueden ser realizados grupalmente.

La entrega de los productos (informe/video) se hace por Comunidades, en un módulo de entregas que se disponibilizará a su debido tiempo. Para la consigna 1 deberá entregarse un archivo .pdf y para la consigna 2 el envío consistirá en un link de acceso público para el que no haga falta ni usuario ni password ni estar registrado en ninguna plataforma para quienes lo vean. Los formatos preferidos para los videos son: YouTube, Prezi o similar. Se desalienta enfáticamente la entrega de links a Google Drive o Dropbox. El video debe ser accesible para la cátedra por lo menos hasta aprobada la asignatura en mesa de examen.

Habrá una nota para el informe a Rubén Feffer y otra nota para el video a Ana Priestly. Ambas notas serán promediadas para definir la nota final del trabajo práctico.

Torneo de Videos

Se propone a lxs estudiantes la realización de un torneo de videos. Mediante este torneo se seleccionará un ganador, quien contará con 1 punto adicional en la nota de la asignatura al momento de ser rendida en mesa de examen (siempre y cuando su examen final esté aprobado).

Los videos se calificarán mediante la metodología Swiss-system tournament con el método de apareo Dutch system. Este método implica la existencia de rondas; en cada ronda hay duelos de videos donde se intenta que compitan videos de similar calidad. Solo se pasa a la ronda siguiente una vez que todos los duelos hayan sido calificados.

Los jueces de cada duelo serán los mismos estudiantes y también lxs docentes. Se espera que cada estudiante sea juez en al menos 4 duelos. Esto significa que cada estudiante deberá ver 8 videos dirigidos a Ana Priestly. Nadie será juez de su propio video ni de los videos de sus compañerxs de grupo. La votación en los duelos de videos es secreta. Se pretende que TODOS los estudiantes participen en todas las rondas como jueces.

Sobre los datos

El Observatorio Villero es una herramienta creada desde La Poderosa en el año 2020. Está conducido por vecinos y vecinas de los barrios populares y su objetivo es visibilizar las condiciones de vida en las villas, asentamientos y barrios populares de Argentina. Las personas que habitan en los barrios son quienes salen a relevar su propia realidad, convirtiéndose en sujetos activos, protagonistas de una historia que hay que poder contar con datos.

En 2022 se tomó la decisión de realizar un Relevamiento de Condiciones Habitacionales, con la colaboración de la Fundación Rosa Luxemburgo. Se diseñó un cuestionario que incorporaba el conocimiento práctico de los territorios habitados, es decir, las variables y las preguntas que mejor permitieran describir la vida cotidiana.

Se relevaron datos de 1222 viviendas en 23 villas y barrios populares de todo el país. Se preguntó sobre las condiciones materiales de las viviendas y el hacinamiento, sobre el acceso al agua y al saneamiento, a la electricidad, al gas, sobre la conectividad, sobre la distribución de espacios verdes y las inundaciones.

Los datos objeto de análisis pueden ser encontrados en este link. Además del acceso a los datos, el link provisto muestra un análisis preliminar de los mismos para cada eje de análisis, con algunas consideraciones y explicaciones de los resultados. Es importante tener en cuenta que no todos los recursos visuales utilizados son completamente correctos desde un punto de vista técnico estadístico, por lo que este análisis puede servir de guía pero deberá considerarse a la luz de los contenidos teóricos provistos por la cátedra de Probabilidad y Estadística.

La tabla de datos que van a utilizar contiene 118 columnas. Se recomienda realizar una primera lectura del contenido publicado en el link, una exploración inicial de los datos

para identificar las variables disponibles, un análisis exploratorio preliminar para ver en términos generales la distribución de las variables y luego una selección posterior del recorte sobre el cual focalizar su análisis.

Las consignas para la generación del informe y video deberán cumplirse teniendo en cuenta las siguiente pautas:

- Puede utilizar recursos gráficos (tablas o gráficos) y/o numéricos, lo que usted considere apropiado para la descripción de los datos.
- Se recomienda fervientemente no presentar la misma información de distintas formas (por ejemplo, una tabla de distribución de frecuencias para una variable cuantitativa continua, el correspondiente histograma y el polígono de frecuencias), sino que es conveniente optar por el recurso que explicite de forma más clara la característica que usted desee mostrar.
- Incluir al menos:
    - descripción gráfica de una variable categórica medida en escala nominal
    - descripción gráfica de una variable categórica medida en escala ordinal
    - descripción gráfica de una variable categórica de respuesta múltiple
    - descripción gráfica de una variable cuantitativa discreta
    - descripción gráfica de una variable cuantitativa continua
    - medida de posición que mejor acompañe a los recursos gráficos anteriormente mencionados
    - consideraciones sobre la dispersión de las variables (puede ser a través de una medida resumen o a partir de algún recurso gráfico), en aquellos casos en los que sea apropiado
    - descripción gráfica de la relación entre dos variables categóricas
    - descripción gráfica de la relación entre una variable categórica y una variable cuantitativa
    - descripción gráfica de la relación entre dos variables cuantitativas

  Este punteo incluye los contenidos que deben incluirse de mínima en las presentaciones, lo que no implica que puedan incluirse más gráficos de un mismo tipo en caso de considerarlo oportuno.
- Seguir las pautas de diseño comunicacional y data storytelling propuestas por la cátedra.

Repositorio: LINK DE GITHUB

Tal como se mencionó previamente, el sector al que usted se suma como data scientist junior cuenta con un ecosistema de repositorios de datos y scripts de análisis. La política de la organización ante miembros de los equipos con menos de un año de antigüedad es la de no otorgar permiso de edición a nada de lo construido. Usted va a poder ver los scripts que se utilizaron con anterioridad en un proyecto de análisis descriptivo de datos. Estos scripts están siendo revisados por uno de sus compañeros Data Vertical, por lo que se pueden ir modificando con el avance de los días/semanas, o incluso pueden agregarse archivos nuevos. Todos ellos se encuentran en este repositorio de GitHub. Usted deberá forkear este repo para contar con los scripts propuestos por la cátedra como punto inicial de su análisis. Recuerde que es una buena práctica sincronizar periódicamente el repositorio forkeado para actualizar los scripts provistos. Se recomienda fervientemente no

modificar los scripts originales sino hacer scripts nuevos que tomen de base los scripts oficiales.

Es parte de la filosofía de la materia la reproducibilidad de los programas. El repositorio de GitHub del grupo deberá permanecer completamente público y abierto, accesible para lxs docentes y sus compañerxs, durante todo el transcurso de la asignatura y hasta adquirida la aprobación de la materia en mesa de exámenes finales.

Ustedes deberán subir a su repositorio grupal de la materia en GitHub los scripts que hayan creado de forma tal que, partiendo del dataset original, permitan a lxs docentes reproducir exactamente los resultados presentados. El link de este repositorio grupal deberá ser informado en la planilla de conformación de grupos. Los grupos deben ser de hasta tres integrantes, y si bien es posible que realice el análisis descriptivo de manera solitaria, es recomendado que conforme un grupo con dos compañerxs más para poder discutir ideas.

Su solución definitiva será analizada minuciosamente por lxs docentes para:

- verificar que las imágenes de las presentación fueron generadas por esos scripts,
- garantizar que está haciendo modificaciones sustantivas a los scripts ofrecidos por la cátedra,
- garantizar que el análisis de datos realizado por su equipo es original y sustancialmente distinto al de los otros grupos.