

“Hackaton de Datos - Potenciando el uso estadístico de los registros administrativos”

Bases del Concurso

El Concurso “Hackaton de Datos - Potenciando el uso estadístico de los registros administrativos” tiene como objetivo principal propiciar el análisis y explotación de los registros administrativos con fines estadísticos y promover espacios para la investigación estadística a través de diferentes herramientas y formatos. Este concurso es organizado por el Instituto Nacional de Estadística (INE), la Agencia de Gobierno Electrónico y Sociedad de la Información y del Conocimiento (AGESIC), el Banco Interamericano de Desarrollo (BID) y el Centro de Pensamiento Estratégico Internacional (CEPEI), los organizadores de aquí en adelante.

El concurso se realiza en el marco del [Plan Estratégico del INE 2020 - 2024](#), el cual incluye entre sus ejes estratégicos y líneas de acción:

- Promover la investigación: Generar y participar en concursos internos y externos para la realización de investigaciones estadísticas.
- Innovar en la producción estadística fomentando el uso de registros administrativos.
- Fortalecimiento de la comunicación con los usuarios, utilizando herramientas de visualización de datos.

1. Objetivos específicos

- Promover el uso, el análisis y la comunicación de los datos en los participantes del concurso.
- Fortalecer las capacidades técnicas de los participantes a través de entrenamientos específicos y el intercambio de conocimiento.
- Fomentar la investigación y generación de estadísticas a partir de registros administrativos.

2. Estructura y duración del Concurso

El Concurso “Hackaton de Datos – Potenciando el uso estadístico de los registros administrativos” se encuentra organizado en cinco fases que incluyen el lanzamiento del concurso, la inscripción, entrenamiento, desarrollo de proyectos y la presentación a los jurados, para finalizar con la premiación de los ganadores.

Las actividades darán comienzo el día 02 de agosto de 2021 con el lanzamiento del presente concurso a través de los canales de comunicación institucional de los organizadores (página web, redes sociales, mails institucionales, convocatorias a las Universidades Públicas y Privadas, notas de prensa, etc.).

Las cinco fases están organizadas de la siguiente manera:

- **Fase I – Lanzamiento del concurso | 02 de agosto:** Publicación de bases del concurso en los canales de comunicación institucional y notas de prensa.

- **Fase II – Inscripción | 02 al 15 de agosto:** Convocatoria a interesados a inscribirse. La inscripción se realizará en dos etapas, una primera en la cual a través del siguiente link (<https://bit.ly/3rF6R4m>) uno de los integrantes inscribe al equipo, indicando nombre del equipo, nombre, apellido y correo electrónico de cada integrante y la categoría en la que se quiere participar. Posteriormente, a cada uno de los integrantes de los equipos que sean seleccionados para participar se le enviará un segundo formulario para completar su información personal y confirmar su participación en la hackatón.
- **Fase III – Entrenamiento | 23 al 27 de agosto:** Fase de carácter opcional en la cual se procederá a la capacitación o actualización en diferentes herramientas y metodologías para analizar y crear las investigaciones.
- **Fase IV – Desarrollo de proyectos | 23 de agosto al 20 de setiembre:** los equipos de trabajo inscritos desarrollan un proyecto de uso de datos.
- **Fase V – Presentación | 23 y 24 de setiembre.** En esta fase se presentará al jurado los productos desarrollados o prototipo de aplicación para calificación según criterios definidos en el numeral 7.1.
- **Entrega de premios | 30 de setiembre:** Durante la jornada de actividades por la celebración del Día de la Estadística Nacional se procederá a la entrega de los premios al equipo ganador de cada categoría de forma presencial, en la medida que la situación coyuntural lo permita y dando cumplimiento a todas las medidas de bioseguridad. No obstante, la decisión final sobre la modalidad del evento se comunicará una semana previa a la fecha de premiación.

No es obligación la participación en las cuatro Fases de este Concurso. Para acceder a la Fase V es requisito excluyente haber participado de las Fase II y IV.

3. Bases específicas para cada Fase

La Fase II de inscripción tiene como objetivo específico fomentar la participación ciudadana mediante la apertura de un espacio público para fortalecer el análisis y explotación de los registros administrativos con fines estadísticos y promover la investigación estadística a través de diferentes herramientas y formatos. Por otro lado también se busca identificar datos que puedan ser de interés de la sociedad en su conjunto, haciendo énfasis en la visualización de los mismos en forma amigable.

Los grupos participantes podrán desarrollar sus trabajos a partir de subconjuntos de datos sintéticos anonimizados, obtenidos a partir de los siguientes conjuntos de datos:

- Registro estadístico de Empresas y Establecimientos (*).
- Registro estadístico de población (*).
- Registros estadísticos de relaciones laborales y remuneraciones (*).
- Registro estadístico de movilidad de celulares (matriz origen-destino).
- Microdatos de la Encuesta Continua de Hogares con descripciones de ocupaciones, ramas de actividad y educación.

(*)Registros estadísticos creados por el INE a partir de la integración de diversas fuentes de registros administrativos. Los tres registros estadísticos pueden vincularse entre sí, a través de claves de identificación (ID estadístico de personas e ID estadístico de empresas).

- Asimismo, se pondrá a disposición imágenes satelitales del país de diferentes períodos, con el objetivo que se puedan desarrollar proyectos de inteligencia artificial para la teledetección de cambios en la cartografía y edificaciones.

Los participantes deberán utilizar los **conjuntos de datos** establecidos en forma **obligatoria** para cada **categoría**. Se les proveerá acceso a los conjuntos de datos basados en “registros estadísticos”, en forma presencial dentro del local del INE (Liniers 1280, 3 y 4 piso). En cambio los microdatos de la ECH e imágenes satelitales podrán descargarse para trabajar offline en forma remota.

Podrán utilizar fuentes adicionales de datos, provistas por cada grupo, siempre y cuando sean fuentes de datos abiertos y cumplan con la normativa vigente en materia de propiedad intelectual.

La Fase III está compuesta por una sesión introductoria y tres sesiones de capacitación o actualización de saberes impartidas por Cepei y el INE en forma virtual. Cada grupo inscrito en el concurso tendrá derecho a un cupo en el taller correspondiente a la categoría inscrita. La participación en capacitaciones diferentes a las relacionadas con la temática seleccionada por el equipo dependerá de la disponibilidad de cupos.

Los temas a abordar en los talleres serán:

- Presentación de los datos disponibles para el concurso y posibles usos de los mismos.
- “Tips” sobre cómo contar historias con datos, haciendo una comunicación efectiva y amigable.
- Herramientas para la visualización de datos y casos de aplicación.
- Presentaciones sobre experiencias internacionales en la aplicación de Machine Learning e Inteligencia Artificial para la codificación automática de textos descriptivos de ocupación, rama de actividad y educación.

El link a cada taller, así como el cronograma de planificación de los mismos serán enviados oportunamente a los participantes a través del correo electrónico establecido en la instancia de inscripción (Fase II).

Las bases específicas de la Fase IV serán publicadas en la página web del INE semanas antes del inicio de la fase.

4. Categorías del concurso

Existen tres categorías para participar del Concurso.

- 1) Contar historias con datos: elaborar historias con datos de alto impacto como un instrumento eficiente, amigable e interactivo para comunicar una problemática, hallazgo o solución basada en datos.
- 2) Visualizaciones: generación de nuevas formas de visualizar y comunicar datos, a fin de incentivar su uso en diferentes ámbitos y temáticas.
- 3) Aplicaciones de Machine Learning o Inteligencia Artificial: crear modelos de machine learning orientados a la codificación automática de textos. Así como también modelos para la teledetección de cambios en edificaciones y cartografía a partir de interpretación de imágenes satelitales del territorio nacional.

Según la categoría, el proyecto deberá cumplir con una serie de condiciones mínimas para ser considerado en la fase V de presentación y su posterior evaluación.

[1. Historias con datos]

- Conjunto de datos a utilizar en forma obligatoria: al menos uno de los registros estadísticos (empresas, población, relaciones laborales) indicados en el apartado 3.
- A partir del día 23 de agosto tendrán la posibilidad de descargar una versión resumida de los conjuntos de datos para que puedan empezar a planificar y diseñar el proyecto. La versión completa de los conjuntos de datos estará disponible para acceder en forma presencial en las instalaciones del INE (Liniers 1280, 3 y 4 piso) durante la semana del 06 al 10 de setiembre en horario nocturno. Durante esa semana se proveerá de un computador para cada equipo con el software y los archivos de datos necesarios para desarrollar el trabajo. Debido a las medidas de prevención contra el Covid-19 se permitirá el acceso de hasta 3 participantes por grupo.
- Toda **historia con datos** debe ser entregada en un documento Word que además del contenido, describa la temática, mensaje clave, público objetivo y relevancia de la historia con datos. Este documento junto con todos los archivos utilizados para desarrollar la historia con datos deben ser subidos a la nube del INE (Sistema Hermes).
- La historia con datos debe ser cargada en la plataforma DataRepublica.org para presentación al jurado. Para esto, a cada grupo le será asignado un usuario y contraseña.
- En caso de que el usuario decida utilizar Tableau, los organizadores le asignarán una licencia temporal Tableau Desktop por dos semanas.
- Herramientas a utilizar: cualquier editor de textos, Tableau, R, QGIS.
- La estructura de la historia con datos debe contar con:
 - **Título** de extensión moderada.
 - **Introducción:** de qué trata la historia, Objetivos de Desarrollo Sostenible (ODS) relacionados con la temática y descripción general de la estructura de la historia
 - **Desarrollo:** contextualización geográfica y temática; capítulos con subtítulos claves que guarden un hilo conductor claro y eviten texto demasiado largos; formatos visuales como: fotos, **visualizaciones**, fotos, vídeos, entre otros.
 - **Conclusiones y recomendaciones** para la temática tratada, resaltar relación con los Objetivos de Desarrollo Sostenible (ODS).

[2. Visualización de datos]

- El taller de visualización de datos se desarrollará a través de Tableau Online. Por lo tanto, los organizadores le asignarán una licencia temporal Tableau Online a los participantes designados por cada grupo para realizar los ejercicios correspondientes a la sesión del taller.
- Conjunto de datos a utilizar en forma obligatoria: al menos uno de los registros estadísticos (empresas, población, relaciones laborales) indicados en el apartado 3.
- A partir del día 23 de agosto tendrán la posibilidad de descargar una versión resumida de los conjuntos de datos para que puedan empezar a planificar y diseñar el proyecto. La versión completa de los conjuntos de datos estará disponible para

acceder en forma presencial en las instalaciones del INE (Liniers 1280, 3 y 4 piso) durante la semana del 06 al 10 de setiembre en horario nocturno. Durante esa semana se proveerá de un computador para cada equipo con el software y los archivos de datos necesarios para desarrollar el trabajo. Debido a las medidas de prevención contra el Covid-19 se permitirá el acceso de hasta 3 participantes por grupo.

- La dinámica de las visualizaciones debe ser de fácil comprensión para cualquier usuario. Por lo cual, se sugiere utilizar iconos o botones que faciliten su interacción.
- Los dashboards deben incluir distintos formatos de gráficas que se encuentren interrelacionadas, a fin de que permitan realizar análisis de múltiples variables en un solo momento.
- Todas las gráficas deben tener títulos claros y concisos, manejar una paleta de colores adecuada, estar intervinculadas e incluir la fuente de los datos.
- Herramientas a utilizar: las visualizaciones pueden ser presentadas en Tableau, R (html generado a partir de paquetes de R), QGIS, siempre y cuando cumpla con los requisitos descritos anteriormente. En caso de que el usuario decida utilizar Tableau, los organizadores le asignarán una licencia temporal Tableau Desktop por dos semanas.
- Los participantes deberán subir todos los archivos correspondientes a la visualización en Tableau o R a la nube del INE (Sistema Hermes).

[3. Machine learning]

- Conjunto de datos a utilizar: al menos una de las fuentes de datos indicadas en el apartado 3.
- Herramientas a utilizar: Python, R, QGIS.
- Los participantes deberán elaborar un documento (Word u Open Office) de mínimo 10 páginas que contenga la siguiente información:
 - **Definición del problema a resolver:** enuncie el problema sobre la temática de su elección y explique cómo piensa resolverlo.
 - **Descripción de las fuentes de datos etiquetadas:** describa cada componente de los datos y etiquetas que se van a clasificar.
 - **Método de procesamiento utilizado:** explique los diferentes procesos previos que realizó para limpiar los datos.
 - **Exploraciones:** presente cómo explora los datos y los hallazgos que considera más relevantes.
 - **Modelo de machine learning investigado:** explique los modelos que ha investigado para resolver el problema y los parámetros utilizados.
 - **Resultados:** presente los resultados del modelado, el rendimiento dada la métrica de evaluación y una prueba realizada con los conjuntos de datos o imágenes proporcionados.
- Todos los archivos utilizados, incluido el código R o Python con comentarios incluidos en el código y el paso a paso, deben ser subidos a la nube del INE (Sistema Hermes).

Entrega de trabajos

Los participantes deberán entregar los trabajos finales en un único archivo comprimido con el nombre del grupo, que deberán subir a la nube del INE (Sistema Hermes). Se deben incluir allí todos los archivos utilizados para desarrollar el proyecto. A modo de ejemplo: documentos de texto, conjuntos de datos utilizados y/o creados, código fuente Python o R

con comentarios incluidos en el código y el paso a paso, libros de trabajo de Tableau, archivos QGIS, entre otros.

El **plazo de entrega final** de los trabajos es el día **20 de setiembre a las 23:59 horas**, a través de la plataforma Hermes del INE.

5. Participación

5.1. ¿A quién está dirigido?

Hackaton de Datos - Potenciando el uso estadístico de los registros administrativos es un concurso abierto a todo público, a excepción de las personas que tengan un vínculo laboral de cualquier naturaleza (contrato directo o a través de un tercero) o como proveedor de servicios, con alguna de las instituciones organizadoras del concurso (INE, AGESIC, BID y CEPEI), con énfasis en: Estudiantes, Emprendedores, Empresas, Activistas, Investigadores, Científicos de datos, Periodistas y Organizaciones públicas y privadas y todas aquellas personas interesadas en aprender/forjar experiencia en la explotación y análisis de datos provenientes de registros administrativos (colectivamente y en adelante los "Concursantes").

5.2 ¿Quién puede participar?

Las personas que cumplan con el proceso de inscripción y con los siguientes criterios de selección:

1. Completar el formulario de inscripción.
2. Residentes en Uruguay con CI uruguaya.
3. No tener vínculo laboral de cualquier naturaleza (contrato directo o a través de un tercero) ni como proveedor de servicios, con alguna de las instituciones organizadoras del concurso (INE, AGESIC, BID, CEPEI).
4. Aceptar los términos y condiciones de la convocatoria.
5. Tener acceso a internet durante el evento.
6. Contar con conocimientos previos según la categoría de interés:
 - a. *Visualización de datos*: conocimiento básico e intermedio de estadísticas y planillas electrónicas, respectivamente, así como conocimiento básico en Tableau o R.
 - b. *Historias con datos*: conocimiento básico en desarrollo sostenible, redacción y gramática intermedia, conocimiento básico de estadística.
 - c. *Machine learning*: conocimientos de nivel intermedio de R o Python.

5.2. ¿Cómo se participa?

Los participantes deben presentarse en un equipo (y sólo uno) de entre 3 y 5 miembros, pudiendo participar en una de las tres categorías establecidas en el concurso.

Para inscribirse en el concurso, el equipo deberá completar un primer formulario en el siguiente link: <https://bit.ly/3rF6R4m> donde se solicita: nombre del equipo, categoría principal y secundaria en cuál desean concursar y nombre, apellido y correo electrónico de cada integrante.

Cada categoría contará con un máximo de 10 equipos concursantes. Si alguna categoría excediera este número de equipos inscritos, se seleccionarán al azar los 10 equipos que pasarán a la fase III. Los equipos que no queden seleccionados en la categoría para la que se inscribieron en primer lugar, podrán participar, sujeto a disponibilidad, en otra categoría de acuerdo con lo expresado oportunamente en el formulario de inscripción.

Los organizadores se reservan el derecho de cancelar alguna de las categorías si no se alcanza un mínimo de 3 grupos inscritos. Estos grupos podrán ser reasignados a la o las categorías que se consideren apropiadas.

5.3. Confirmación de inscripción

Una vez cerrado el período de inscripción, en base a los datos vertidos en el formulario de inscripción se enviará a cada integrante de los equipos un formulario con el cual se le solicitará la confirmación de participación del concurso, consultándose además sobre su nivel de formación y el o los talleres de entrenamiento a los que le interesa asistir. Este formulario deberá ser completado y enviado antes de las 23:59hs del día 18 de agosto de 2021. De no contarse con dicha confirmación, los organizadores se reservan el derecho de convocar de la lista de inscriptos a tantos grupos como sean necesarios hasta completar los 10 grupos por categoría.

6. Jurado

El proceso de evaluación de los trabajos presentados estará a cargo de un panel de cinco jueces, conformado por representantes de las instituciones organizadoras del concurso: INE, AGESIC, BID y CEPEI, más un representante del sector académico como jurado externo. Los jueces evaluarán el trabajo de cada grupo inmediatamente después de su presentación, de acuerdo con los criterios de valoración y escala de puntuación definidos en el numeral 7.1

Las responsabilidades de los jurados son:

- Evaluar cada postulación de acuerdo con los “Criterios de valoración”.
- Otorgar o dejar vacantes los Premios correspondientes a todas y cada una de las “Categorías de premio”.
- Otorgar, si resultara oportuno, menciones especiales a aquellas postulaciones que sin haber obtenido un Premio, acrediten los méritos suficientes.

Las decisiones del Jurado son definitivas e inapelables.

7. Criterios de valoración de trabajos presentados

Cada jurado evaluará cada proyecto otorgando un máximo de 40 puntos a partir de la suma del puntaje asignado a los conceptos a valorar (Ver numeral 7.1), con una puntuación mínima de 0 y máxima de 10 en cada uno de ellos. El puntaje final corresponderá al promedio de la calificación final de los jueces.

En caso de empate, se seleccionará al equipo que tenga el mayor puntaje en el concepto de innovación, seguido por el mayor puntaje en la relevancia, seguido por el criterio de

investigación o parametrización para la reutilización (según la categoría), hasta que se rompa el empate y solo exista un único ganador. En caso de un empate en todos los criterios, se distribuirá el premio entre los dos ganadores.

7.1. Contar historias con datos

Concepto a valorar	Descripción	Escala (puntos)
Nivel de implementación	Grado de completitud de la propuesta al finalizar la hackatón	1 a 10
Innovación	Carácter innovador de la propuesta	1 a 10
Relevancia	Desarrollo de una temática importante o necesaria para el contexto nacional o territorial basada en datos	1 a 10
Investigación	Análisis del contexto para vincular los datos con la realidad territorial o poblacional que presenta la historia con datos y generación de hipótesis o inferencias a partir del análisis cuantitativo.	1 a 10

7.2. Visualizaciones:

Concepto a valorar	Descripción	Escala (puntos)
Nivel de implementación	Grado de completitud de la propuesta al finalizar la hackatón	1 a 10
Innovación	Carácter innovador de la propuesta	1 a 10
Relevancia	Desarrollo de una temática importante o necesaria para el contexto nacional o territorial basada en datos	1 a 10
Investigación	Recopilación, procesamiento y análisis de datos oportunos para construir la visualización	1 a 10

7.3. Aplicaciones de Machine Learning e Inteligencia Artificial:

Concepto a valorar	Descripción	Escala (puntos)
--------------------	-------------	-----------------

Nivel de implementación	Grado de completitud de la propuesta al finalizar la hackatón	1 a 10
Innovación	Carácter innovador de la propuesta	1 a 10
Relevancia	Desarrollo de una temática importante o necesaria para el contexto nacional o territorial basada en datos	1 a 10
Parametrización para la reutilización	Grado de parametrización de la aplicación que facilita su reutilización en diferentes contextos sin necesidad de modificar el código fuente.	1 a 10

8. Premios

Se premiará al primer puesto de cada una de las categorías de acuerdo con los criterios de valoración, siempre y cuando haya alcanzado un mínimo de 28 puntos.

El equipo ganador de cada categoría se anunciará el 28 de septiembre una vez se consoliden las evaluaciones de todos los jurados y el día 30 de septiembre deberán hacer una presentación de los trabajos realizados en la Hackatón, en el marco del evento de conmemoración del Día de la Estadística Nacional, oportunidad en la que se les hará entrega de los certificados correspondientes.

El ganador tendrá un plazo de siete (7) días calendario para aceptar el premio, a partir de la comunicación de los ganadores en el sitio web del concurso y/o en las redes sociales antes mencionadas. En caso contrario se entregarán por única vez al siguiente equipo participante con mayor puntaje obtenido.

Los beneficios de participar son:

- Todos los participantes recibirán un certificado que acredita su participación y aporte a la hackathon.
- Al equipo ganador de cada categoría se le otorgará un premio de 45.000 pesos uruguayos.
- Cada integrante del equipo ganador de cada categoría recibirá un certificado que acredita el logro alcanzado.
- Publicar los trabajos en la página web del INE, DataRepública- Cepei y AGESIC.

9. Consultas y Notificaciones

Se recibirán consultas sobre este Concurso a la casilla de correo electrónico hackaton2021@ine.gub.uy.

Los concursantes serán notificados de todos los actos, resultados e información general sobre el presente Concurso, a través del sitio web <https://www3.ine.gub.uy/Hackaton/index.html> y redes sociales **@ineuruguay** en Instagram y **@ine_uruguay** en Twitter. Es responsabilidad de los mismos informarse a través de estos medios.

9.1. Aceptación

Por el solo hecho de participar en cualquiera de las Fases de este Concurso, los concursantes declaran que conocen y aceptan en su totalidad las presentes bases.

Cualquier instancia no prevista por las presentes bases, serán resueltas por los organizadores.

9.2. Propiedad intelectual

Los concursantes serán los exclusivos responsables del cumplimiento de la normativa vigente en materia de propiedad intelectual. Los organizadores se exoneran de cualquier tipo de responsabilidad por la vulneración de dicha normativa. Si algún participante o interesado conoce una conducta lesiva a estos derechos podrá comunicarlo a través del correo electrónico hackaton2021@ine.gub.uy.

Por el solo hecho de presentarse los concursantes:

- a) Entienden que de acuerdo a la normativa nacional en propiedad intelectual (Ley N° 9.739 en la redacción dada por la Ley N°17.616, sus modificativas y decreto reglamentario), las ideas presentadas y las propuestas desarrolladas en el concurso no se encuentran protegidas por el derecho de autor.
- b) Los productos y documentos, de cualquier tipo, desarrollados en el marco del presente concurso podrán ser utilizados por el INE a los fines que mejor entienda incluyendo su divulgación, cesión a terceros o utilización para fines de investigación entre otras finalidades.

9.3. Secreto estadístico

El manejo de la información de este concurso (incluidos los conjuntos de datos que estarán a disposición de los participantes según se indica en el apartado “3. Bases específicas para cada fase”), se gestionará bajo lo dispuesto en el artículo 17 de la Ley N° 16.616, de 20 de octubre de 1994.

“Están amparados por el secreto estadístico los datos individuales proporcionados a los organismos del Sistema Estadístico Nacional por las fuentes de información. La obligación de guardar el secreto estadístico alcanza tanto a los organismos como a sus funcionarios, así como a terceras personas que tomen conocimiento de los datos relevados al amparo del mismo”.

Los datos anonimizados a los cuales accederán los participantes podrán ser utilizados exclusivamente a los fines de este concurso y no podrán ser transferidos o divulgados a terceros, ni reproducidos por ningún medio.

El artículo 30 de la precitada Ley 16.616 establece que “las personas que violen el Secreto Estadístico o utilicen los datos individuales en provecho propio o ajeno serán penalmente responsables por el delito tipificado en el artículo 301 del Código Penal “Revelación de documentos secretos”, sin perjuicio de las reparaciones de orden civil que correspondieren”.