

Instituto Tecnológico de Costa Rica

Ingeniería en Computación

Sede Cartago

Inteligencia Artificial

Proyecto corto 1: Generación de muestras para  
simular los votos de la primera ronda electoral.

Profesor:

Juan Esquivel.

Estudiantes:

María Fernanda Alvarado Vargas - 2014084573

Freyser Jiménez Mena - 2014160018

Minor Sancho Valverde – 2014047710

Fecha de entrega:

25 marzo 2018

## Tabla de contenido

<b>Tabla de contenido</b>	<b>1</b>
<b>1. Introducción.</b>	<b>3</b>
<b>2. Descripción del problema.</b>	<b>4</b>
Módulo simulador	4
Fuentes de datos	5
Detalles de Entrega y Revisión	6
<b>3. Análisis del problema.</b>	<b>7</b>
<b>3.1 Análisis de muestras nacionales.</b>	<b>7</b>
3.1.1 Análisis de votos por partido.	7
3.1.2 Análisis de intención de voto por edad.	9
3.1.3 Análisis de intención de voto por partido según edad.	10
3.1.4 Análisis de voto por sexo.	11
<b>3.2 Análisis de muestras por provincia.</b>	<b>12</b>
<b>3.2.1 San José.</b>	<b>12</b>
3.2.1.1 Análisis probabilidad de voto por partido.	12
3.2.1.3 Análisis de intención de voto por partido según la edad.	15
3.2.1.4 Análisis de voto por sexo.	16
<b>3.2.2 Alajuela.</b>	<b>17</b>
3.2.2.1 Análisis probabilidad de voto por partido.	17
3.2.2.2 Análisis de intención de voto por edad.	19
3.2.2.3 Análisis de intención de voto por partido según la edad.	20
3.2.2.4 Análisis de voto por sexo.	21
<b>3.2.3 Cartago.</b>	<b>22</b>
3.2.3.1 Análisis probabilidad de voto por partido.	22
3.2.3.2 Análisis de intención de voto edad.	24
3.2.3.3 Análisis de intención de voto por partido según la edad.	25
3.2.2.4 Análisis de voto por sexo.	27
<b>3.2.4 Heredia.</b>	<b>28</b>
3.2.4.1 Análisis probabilidad de voto por partido.	28
3.2.4.2 Análisis de intención de voto por edad.	30
3.2.4.3 Análisis de intención de voto por partido según la edad.	31
3.2.4.4 Análisis de voto por sexo.	32
<b>3.2.5 Guanacaste.</b>	<b>33</b>
3.2.5.1 Análisis por partido Análisis probabilidad de voto por partido.	33
3.2.5.2 Análisis de intención de voto por edad.	35

3.2.5.3 Análisis de intención de voto por partido según la edad.	36
3.2.5.4 Análisis de voto por sexo.	38
3.2.6 Puntarenas.	39
3.2.6.1 Análisis probabilidad de voto por partido.	39
3.2.6.2 Análisis de intención de voto por edad.	41
3.2.3.6 Análisis de intención de voto por partido según la edad.	42
3.2.6.4 Análisis de voto por sexo.	44
3.2.7 Limón.	44
3.2.7.1 Análisis probabilidad de voto por partido.	44
3.2.7.2 Análisis de intención de voto por edad.	46
3.2.7.3 Análisis de intención de voto por partido según la edad.	47
3.2.7.4 Análisis de voto por sexo.	49
4. Conclusión.	50

## 1. Introducción.

---

Inteligencia Artificial se entiende como la capacidad que tiene una máquina para aprender un entorno; para que esto sea posible se debe obtener un conjunto de datos que sirvan de muestras para lograr entrenar un modelo y de esta forma predecir lo que puede suceder en un ambiente. Este proyecto corto tiene como objetivo generar las muestras necesarias para utilizarlas en un proyecto más grande.

Debido a la situación actual que se está viviendo en el país; las muestras que se van a generar están relacionadas con los votos de la primera ronda electoral; cada uno de los datos que se vayan a utilizar (densidad poblacional, porcentaje de la población analfabeta por edades, porcentaje de personas desempleadas, cantidad de votos por: cantón, partido) en el proyecto serán extraídos de las bases de datos públicas del Estado de la Nación y el Tribunal Supremo de Elecciones. Las muestras generadas se distribuirán de dos formas:

- ❖ País.
- ❖ Provincia.

## 2. Descripción del problema.

---

Un problema significativo al desarrollar agentes de inteligencia artificial radica en la poca disponibilidad de conjuntos de datos de ejemplo. Si bien es cierto existen proyectos abiertos que dan acceso a datos en ciertos dominios, la aplicabilidad de estos es debatible y no necesariamente reflejan casos de la vida real.

Una manera de evitar este problema es generar datos sintéticos. Esto es, asumiendo que sabemos la distribución teórica de los datos, crear programas que obtengan muestras extraídas de la distribución para emular la población.

En este proyecto se pide crear un módulo de Python que contiene funciones para crear muestras de ejemplo, basados en los resultados electorales presidenciales de la primera ronda en Costa Rica.

### Módulo simulador

Se pide a los estudiantes crear dos funciones primarias:

- ❖ **generar\_muestra\_pais(n):** retornará una muestra simulada con n votantes con etiquetas y atributos proporcionales a la distribución de resultados nacionales.
- ❖ **generar\_muestra\_provincia(n, nombre\_provincia):** retornará una muestra simulada siguiendo la distribución de datos para una provincia específica (donde nombre\_provincia es uno de "SAN JOSE", "ALAJUELA", "CARTAGO", "HEREDIA", "GUANACASTE", "PUNTARENAS", "LIMON")

El resultado de llamar a cada función será una lista de listas de Python. El tamaño de la lista externa será n, mientras que cada lista interna tendrá [numero\_atributos + 1] elementos. Los atributos corresponde a características cantonales recolectadas en el censo 2011. La etiqueta, que es la columna adicional a los atributos, será la representación textual de un partido político (nombre en mayúscula y de la manera escrita en las actas de escrutinio) o bien las palabras "NULO" o "BLANCO" cuando corresponda.

Para generar las muestras se deberá seguir el algoritmo básico aleatorizado descrito en clase (favor no usar otro algoritmo sin consultar con el profesor). El objetivo básico de

dicho algoritmo es representar la masa poblacional de cada cantón y las preferencias de voto a este nivel geográfico, procurando que las muestras generadas sean fieles a estas distribuciones.

### Fuentes de datos

Existen 3 fuentes de datos primarias:

- ❖ Actas de escrutinio: Contienen los totales de votaciones por junta receptora y se pueden descargar a través de [http://www.tse.go.cr/elecciones2018/actas\\_escrutinio.htm](http://www.tse.go.cr/elecciones2018/actas_escrutinio.htm)
- ❖ Mapeo juntas a cantones: Dado que los resultados de las juntas no contienen explícitamente el cantón, tendremos que reproducirlo con la información contenida en: <http://www.tse.go.cr/pdf/nacional2018/JRV.pdf>
- ❖ Indicadores cantonales: Corresponden a los atributos que nos permitirán generar predicciones para cada uno de los ejemplos. Estos se extraen del Censo 2011 y se encuentran almacenadas en [https://www.estadonacion.or.cr/files/biblioteca\\_virtual/otras\\_publicaciones/Indicadores-cantoniales\\_Censos-2000-y-2011.xlsx](https://www.estadonacion.or.cr/files/biblioteca_virtual/otras_publicaciones/Indicadores-cantoniales_Censos-2000-y-2011.xlsx)

Los estudiantes estarán a cargo de curar esa información, y generar una serie de archivos CSV, que puede usarse como entrada al módulo. Entregables:

- ❖ Código fuente de las dos funciones mencionadas contenidas en un módulo llamado `tec.ic.ia.pc1.g##`, con `pc1` representado Proyecto Corto 1, y `g##` que se reemplazará por el número de grupo asignado por el profesor. Esto quiere decir, por ejemplo, que una vez entregada la tarea podríamos tener la siguiente cláusula de importación para llamar a las funciones:

```
from tec.ic.ia.pc1.g01 import generar_muestra_pais, generar_muestra_provincia
```

Lo anterior será vital ya que parte de la revisión será realizada asumiendo que se pueden importar las funciones de esa forma. Nótese que esto implica que estructura de carpetas debe ser esa en específico.

- ❖ Los archivos CSV utilizados por las funciones, deben ser agregados al paquete enviado. Estos mismos archivos deberán ser leído por el código desarrollado. Crear variables globales al módulo que nos permitan cambiarlos fácilmente, es un requerimiento para todos los archivos (idealmente usamos parámetros a la consola, pero se dejará este nivel para el proyecto).
- ❖ Pruebas unitarias. Si bien es cierto no se espera que haya pruebas que revisen con rigurosidad estadística si las muestra simuladas en efecto vienen de la distribución de los datos, si se pide que haya pruebas que puedan revisar las funcionalidades. Por ejemplo, los estudiantes podrían "inyectar" un generador de números aleatorios para pruebas, cuyos valores son determinísticos, permitiéndonos escribir pruebas. Las mismas deben realizarse utilizando pytest.
- ❖ Un informe, en formato PDF, sobre la comparación de resultados obtenidos durante múltiples ejecuciones del simulador. Si todo estuviera bien, al correr las funciones muchas veces debería crear, en promedio, muestras que aproximen con relativa certeza. Debe mostrarse análisis nacionales y para cada provincia. El nivel de detalle del análisis tendrá un impacto directo en la calificación de este rubro.

### Detalles de Entrega y Revisión

El proyecto corto deberá realizarse en grupos de 3 personas. Se debe enviar todos los entregables a través de TEC Digital a más tardar el 25 de marzo del año en curso.

El profesor se reserva el derecho a asignar una nota de cero si los requerimientos no son cumplidos de forma tal que impida ejecutar las funciones principales.

### Consideraciones Python

- ❖ Se utilizará Python 3 (no 2.7)
- ❖ Se utilizará pip para instalar el módulo enviado. Se debe respetar la estructura de directorios apropiada para que la instalación sea exitosa.
- ❖ Se debe utilizar pytest para pruebas unitarias
- ❖ Seguir recomendaciones en <http://docs.python-guide.org/en/latest/>
- ❖ Seguir PEP 20. Una guía rápida con ejemplos está disponible en: [http://artifex.org/~hblanks/talks/2011/pep20\\_by\\_example.pdf](http://artifex.org/~hblanks/talks/2011/pep20_by_example.pdf)

- ❖ Utilizar PEP 8 como guía de estilo: <http://pep8.org/>

### 3. Análisis del problema.

En esta sección se muestra un análisis de los datos generados tanto nacionalmente como por cada provincia; para verificar que los datos se aproximen a la realidad, se van a realizar varias pruebas y se van a comparar con los datos reales.

#### 3.1 Análisis de muestras nacionales.

Para realizar este análisis se utilizó una muestra de 10 000 personas.

##### 3.1.1 Análisis de votos por partido.

Se realiza una función que estime la cantidad de votos por cada provincia para cada uno de los partidos disponibles, tomando en cuenta los votos nulos y los votos blancos.

#### Datos utilizados:

A continuación, se muestra una tabla con las probabilidades descritas anteriormente.

Partido	San José	Alajuela	Cartago	Heredia	Guanacaste	Puntarenas	Limón
ACCESIBILIDAD SIN EXCLUSIÓN	11	2	3	2	1	1	3
ACCION CIUDADANA	843	441	328	286	94	90	79
ALIANZA DEMOCRATICA CRISTIANA	11	6	31	5	0	2	5
DE LOS TRABAJADORES	9	3	0	1	0	0	2
FRENTE AMPLIO	19	15	9	8	2	6	6



INTEGRACION NACIONAL	314	183	134	105	85	88	83
LIBERACION NACIONAL	620	343	244	187	169	157	126
MOVIMIENTO LIBERTARIO	39	17	11	14	2	7	7
NUEVA GENERACION	37	9	14	8	4	6	2
RENOVACION COSTARRICENSE	24	10	7	4	5	5	4
REPUBLICANO SOCIAL CRISTIANO	154	87	79	40	25	32	27
RESTAURACION NACIONAL	834	513	179	232	152	229	301
UNIDAD SOCIAL CRISTIANA	642	286	209	183	107	118	50
VOTOS NULOS	20	15	17	5	14	13	12
VOTOS BLANCOS	6	6	3	1	0	1	7

Gráfico:

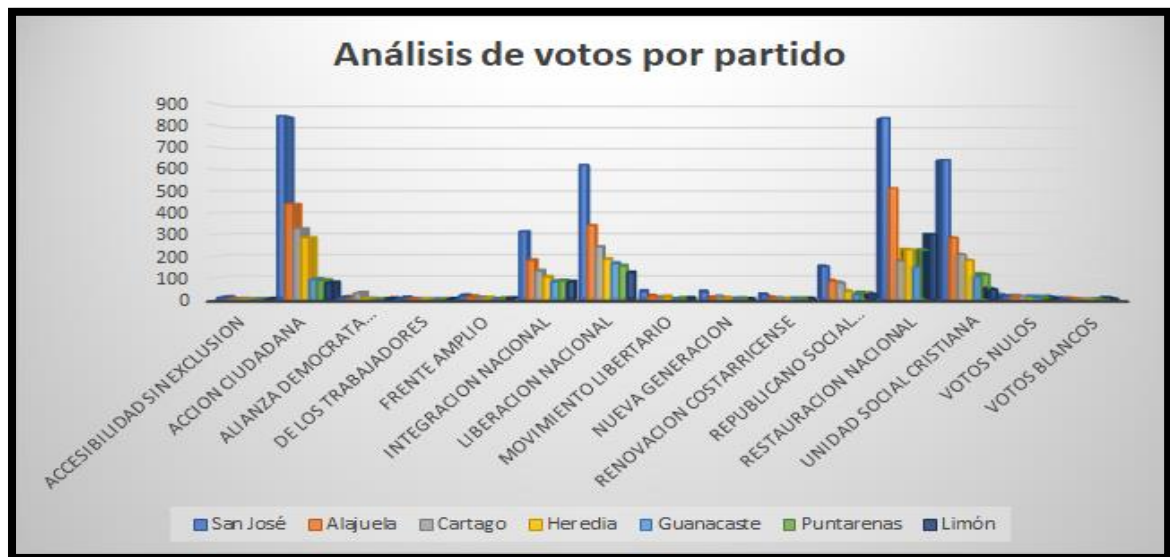


Figura 1: Gráfico que muestra la distribución de los votos de cada partido por provincia.

### 3.1.2 Análisis de intención de voto por edad.

- ❖ Probabilidad de voto estimado: Se realiza una función que estima una probabilidad de votantes que están en un rango de edades. Esto se hace a nivel nacional.

#### Datos utilizados:

A continuación, se muestra una tabla con las probabilidades descritas anteriormente.

Edad	Probabilidad de votos por edad
18-25 años	0,125
26-40 años	0,2365
41-55 años	0,2308
55 y más años	0,4077

Gráfico:

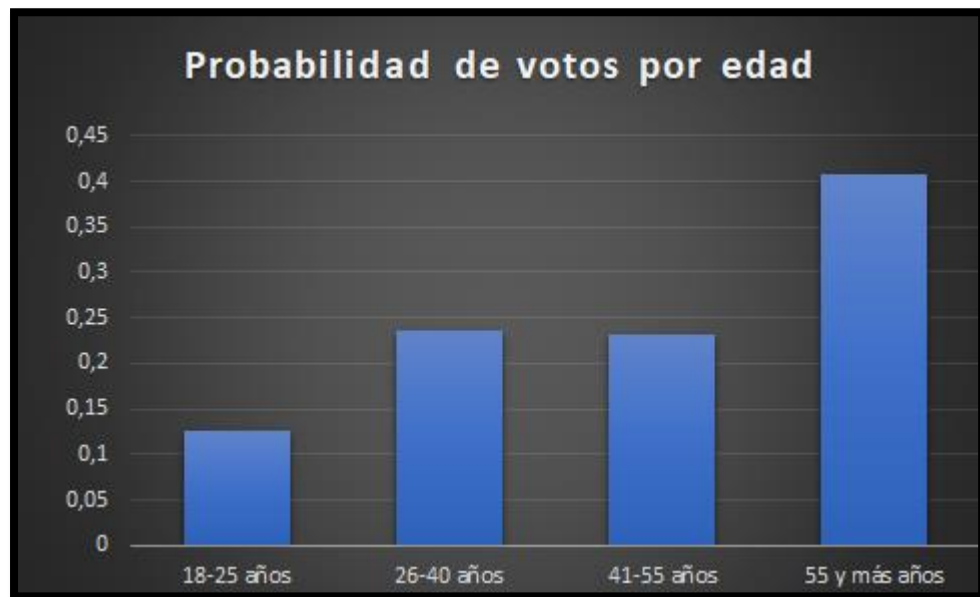


Figura 2: Gráfico que muestra la distribución de los votos según los diferentes rangos de edades.

### 3.1.3 Análisis de intención de voto por partido según edad.

En este análisis se pretende observar la estimación de votos que tiene cada partido según el rango de edades descritas en el análisis anterior.

Datos utilizados:

Partido	18-25 años	26-40 años	41-55 años	55 y más años
ACCESIBILIDAD SIN EXCLUSION	0,0005	0,0005	0,0009	0,0011
ACCION CIUDADANA	0,0268	0,0501	0,0479	0,089
ALIANZA DEMOCRATA CRISTIANA	0,0009	0,0021	0,0016	0,0014
DE LOS TRABAJADORES	0,0003	0,0004	0,0005	0,0007
FRENTE AMPLIO	0,0006	0,0015	0,0015	0,0036
INTEGRACION NACIONAL	0,0122	0,023	0,02	0,038
LIBERACION NACIONAL	0,0235	0,044	0,0443	0,0724
MOVIMIENTO LIBERTARIO	0,0012	0,0025	0,0023	0,0032
NUEVA GENERACION	0,0009	0,0017	0,0014	0,004

RENOVACION COSTARRICENSE		0,0007	0,0013	0,0019	0,0023
REPUBLICANO SOCIAL CRISTIANO		0,0047	0,0114	0,0115	0,0187
RESTAURACION NACIONAL		0,0305	0,0594	0,0583	0,1001
UNIDAD SOCIAL CRISTIANA		0,0214	0,0414	0,0378	0,0619
VOTOS NULOS		0,0007	0,0026	0,0026	0,0025
VOTOS BLANCOS		0,0005	0,0002	0,0004	0,0007

Gráfico:



Figura 3: Gráfico que muestra la distribución de los votos por cada partido según los rangos de edad mostrados.

### 3.1.4 Análisis de voto por sexo.

Datos utilizados:

Provincia	Hombre	Mujer
San José	2342	1208
Alajuela	1406	573
Cartago	877	389
Heredia	683	362
Guanacaste	483	206
Puntarena	588	195

s		
Limón	506	182

Gráfico:

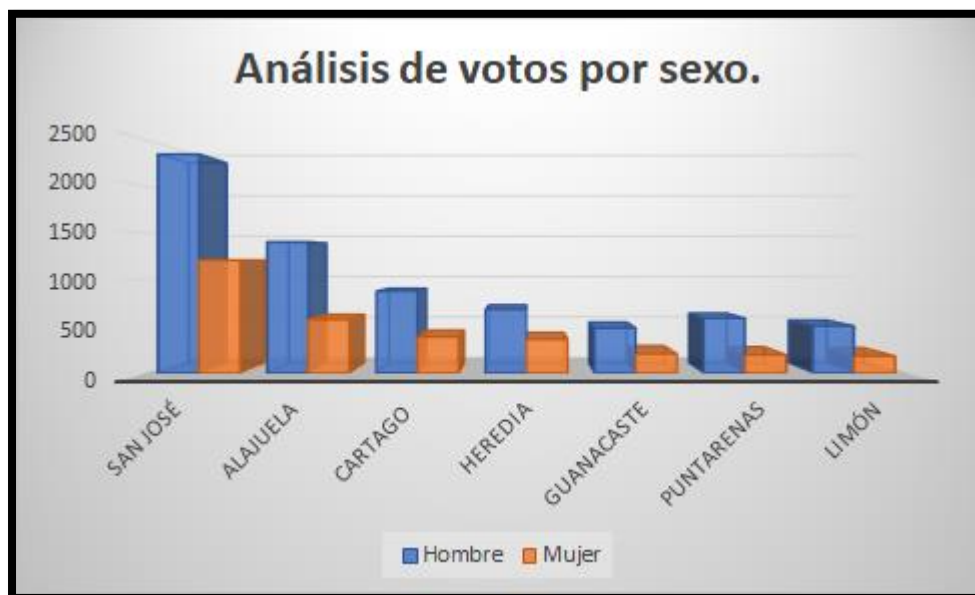


Figura 4: Gráfico que muestra la distribución de los votos por cada partido según el sexo de las personas; se puede notar que la mayoría de los votos son emitidos por hombres.

### 3.2 Análisis de muestras por provincia.

#### 3.2.1 San José.

##### 3.2.1.1 Análisis probabilidad de voto por partido.

- ❖ Probabilidad de voto real: Son las probabilidades reales basadas en los datos que el profesor brindó.
- ❖ Probabilidad de voto estimado: Para estimar esta probabilidad se implementó una función que simula 5 veces los datos y las probabilidades que dicha función retorna una vez estimados estos datos se promedian y se obtiene la probabilidad deseada. Es importante recalcar que las pruebas realizadas se hacen con una muestra de 5000 personas.

#### Datos utilizados:

A continuación, se muestra una tabla con las probabilidades descritas anteriormente.

Partidos políticos	Probabilidad de voto real	Probabilidad de voto simulada
ACCESIBILIDAD SIN EXCLUSION	0.001	0.001
ACCION CIUDADANA	0.082	0.083
ALIANZA DEMOCRATA CRISTIANA	0.001	0.001
DE LOS TRABAJADORES	0.0006	0.0006
FRENTE AMPLIO	0.002	0.002
INTEGRACION NACIONAL	0.031	0.03
LIBERACION NACIONAL	0.063	0.064
MOVIMIENTO LIBERTARIO	0.003	0.003
NUEVA GENERACION	0.002	0.002
RENOVACION COSTARRICENSE	0.001	0.001
REPUBLICANO SOCIAL CRISTIANO	0.016	0.016
RESTAURACION NACIONAL	0.08	0.08
UNIDAD SOCIAL CRISTIANA	0.062	0.062
VOTOS NULOS	0.002	0.002
VOTOS BLANCOS	0.0005	0.0005

## Gráfico:

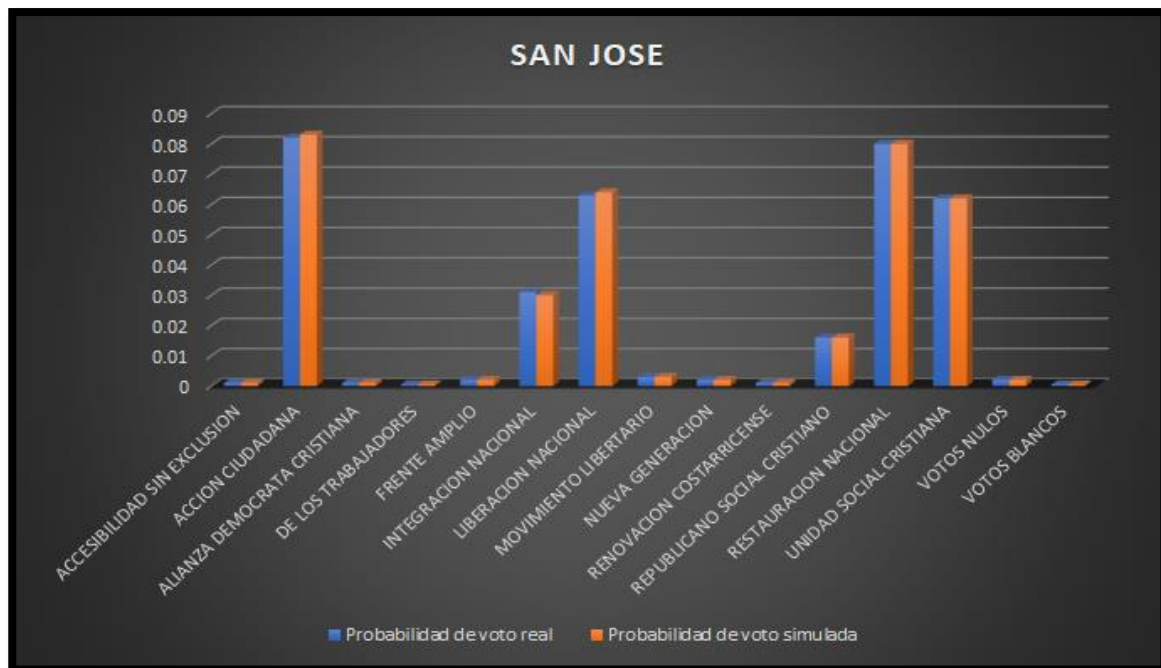


Figura 5: Gráfico que compara la probabilidad de votos reales emitidos y la probabilidad estimada por cada partido; este gráfico se obtiene según los datos mostrados anteriormente. Además, se puede visualizar que los datos no varían mucho y las diferencias que se dan no son tan significativas; es decir, no se alteran demasiado.

### 3.2.1.2 Análisis de intención de voto por edad.

- ❖ Probabilidad de voto estimado: Se realiza una función que estima una probabilidad de votantes que están en un rango de edades.

#### Datos utilizados:

A continuación, se muestra una tabla con las probabilidades descritas anteriormente.

Edades	Probabilidad de voto simulada
18 a 25 años (Jóvenes)	0.1226
26 a 40 años (Adulto joven)	0.2494
41 a 55 años (Adultos)	0.2202
56 o más (Adulto mayor)	0.4078

Gráfico:

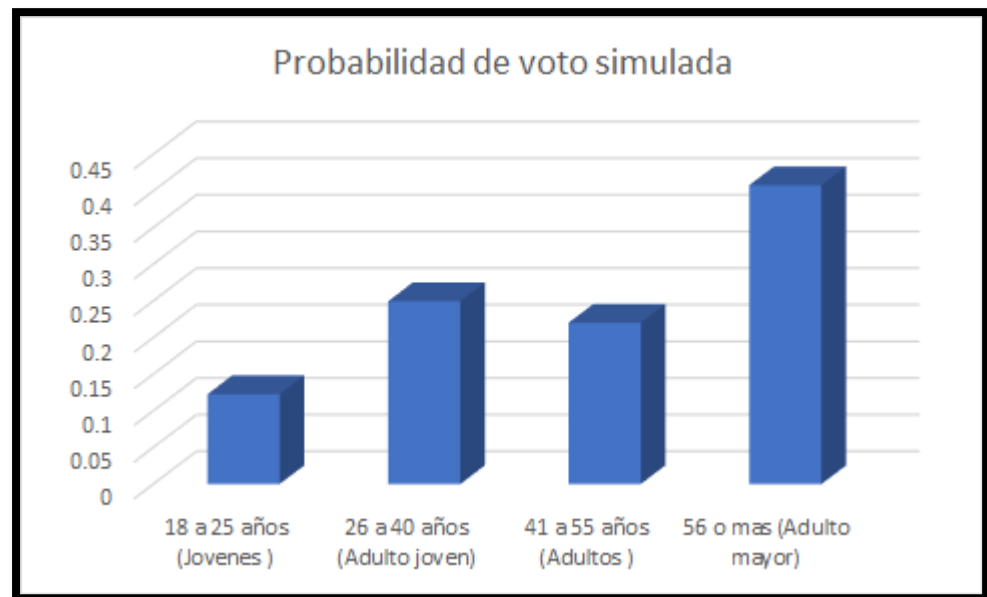


Figura 6: El gráfico muestra que, en la simulación de votos realizada, la mayoría de la población que ejerce el voto están en un rango de 26 a 40 años y de 56 años en adelante.

### 3.2.1.3 Análisis de intención de voto por partido según la edad.

En este análisis se pretende observar la estimación de votos que tiene cada partido según el rango de edades descritas en el análisis anterior.

**Datos utilizados:**

Partidos políticos	18 a 25 años (Jóvenes)	26 a 40 años (Adulto joven)	41 a 55 años (Adultos)	56 o más (Adulto mayor)
ACCESIBILIDAD SIN EXCLUSION	1	0	5	10
ACCION CIUDADANA	138	268	268	477
ALIANZA DEMOCRATA CRISTIANA	2	3	4	11
DE LOS TRABAJADORES	2	2	1	4
FRENTE AMPLIO	4	10	5	17
INTEGRACION NACIONAL	48	103	101	195
LIBERACION NACIONAL	106	242	179	385



MOVIMIENTO LIBERTARIO	10	15	9	24
NUEVA GENERACION	8	8	14	16
RENOVACION COSTARRICENSE	5	10	5	15
REPUBLICANO SOCIAL CRISTIANO	32	60	48	88
RESTAURACION NACIONAL	136	326	221	424
UNIDAD SOCIAL CRISTIANA	117	233	198	338
VOTOS NULOS	5	13	10	15
VOTOS BLANCOS	2	1	2	1
VOTOS TOTALES	616	294	1070	2020

### Gráfico.

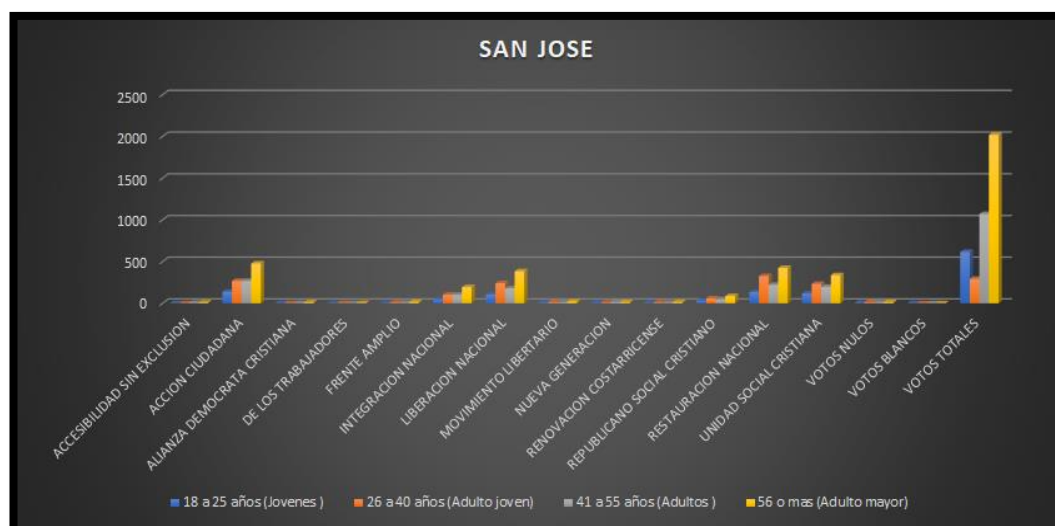


Figura 7: Gráfico que muestra una simulación de votos por partido según un rango de edad.

#### 3.2.1.4 Análisis de voto por sexo.

Datos utilizados:

Genero	Probabilidad de voto simulado
Hombres	0.6578
Mujeres	0.3422

Gráfico:

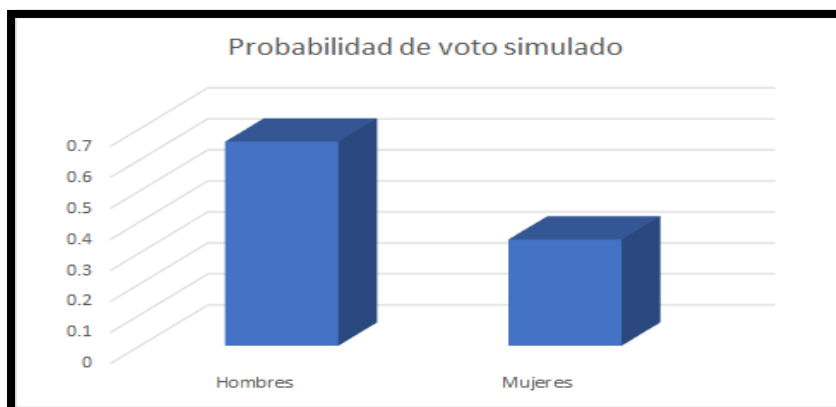


Figura 8: Gráfico que muestra una simulación de votos por sexo de las personas.

### 3.2.2 Alajuela.

#### 3.2.2.1 Análisis probabilidad de voto por partido.

- ❖ Probabilidad de voto real: Son las probabilidades reales basadas en los datos que el profesor brindó.
- ❖ Probabilidad de voto estimado: Para estimar esta probabilidad se implementó una función que simula 5 veces los datos y las probabilidades que dicha función retorna una vez estimados estos datos se promedian y se obtiene la probabilidad deseada. Es importante recalcar que las pruebas realizadas se hacen con una muestra de 5000 personas.

#### Datos utilizados:

A continuación, se muestra una tabla con las probabilidades descritas anteriormente.

Partidos políticos	Probabilidad de voto real	Probabilidad de voto simulada
ACCESIBILIDAD SIN EXCLUSION	0.0006	0.0006

ACCION CIUDADANA	0.04	0.04
ALIANZA DEMOCRATA CRISTIANA	0.0006	0.0006
DE LOS TRABAJADORES	0.0004	0.0005
FRENTE AMPLIO	0.001	0.001
INTEGRACION NACIONAL	0.01	0.01
LIBERACION NACIONAL	0.03	0.03
MOVIMIENTO LIBERTARIO	0.002	0.002
NUEVA GENERACION	0.001	0.001
RENOVACION COSTARRICENSE	0.0009	0.0004
REPUBLICANO SOCIAL CRISTIANO	0.01	0.01
RESTAURACION NACIONAL	0.05	0.05
UNIDAD SOCIAL CRISTIANA	0.03	0.03
VOTOS NULOS	0.001	0.001
VOTOS BLANCOS	0.0004	0.0002

Gráfico:

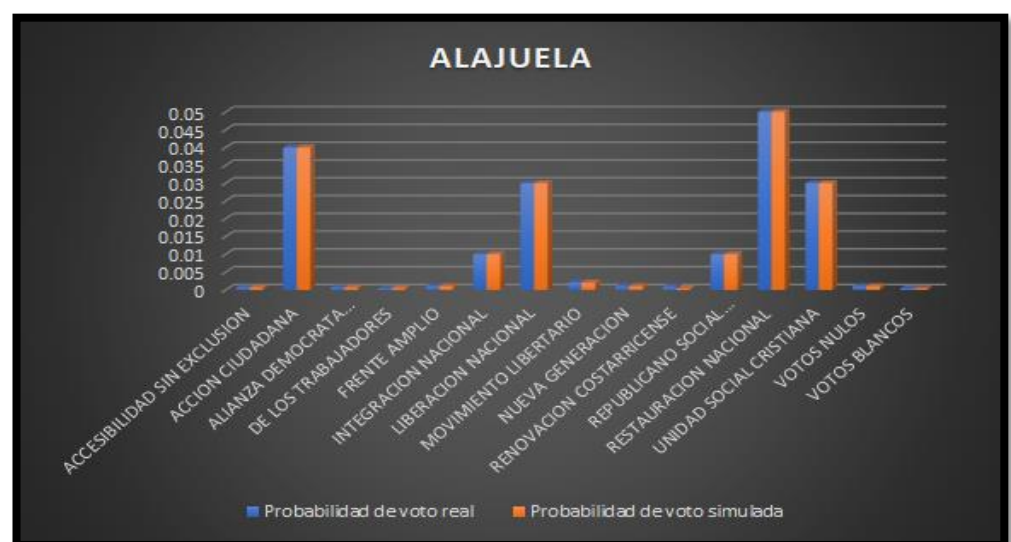


Figura 9: El gráfico mostrado hace una comparación de la probabilidad de votos reales emitidos y la probabilidad estimada por cada partido; este gráfico se obtiene según los datos mostrados anteriormente. Además, se puede visualizar que los datos no varían mucho y las diferencias que se dan no son tan significantes; es decir, no se alteran demasiado.

### 3.2.2.2 Análisis de intención de voto por edad.

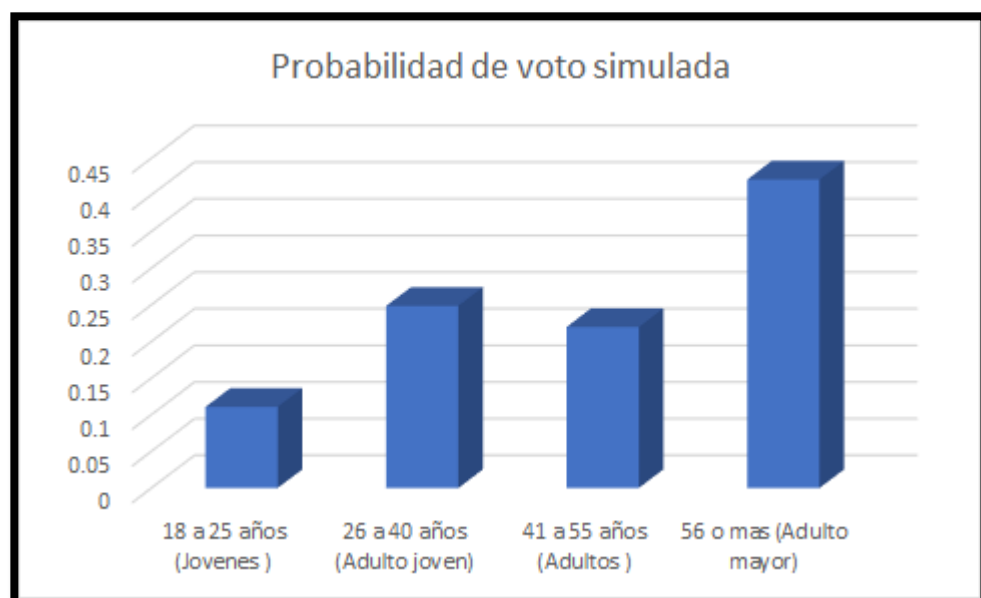
- ❖ Probabilidad de voto estimado: Se realiza una función que estima una probabilidad de votantes que están en un rango de edades entre 18 a 24 años y de 25 años en adelante.

#### Datos utilizados:

A continuación, se muestra una tabla con las probabilidades descritas anteriormente.

Edades	Probabilidad de voto simulada
18 a 25 años (Jóvenes)	0.1108
26 a 40 años (Adulto joven)	0.2484
41 a 55 años (Adultos)	0.2198
56 o más (Adulto mayor)	0.421

#### Gráfico:



*Figura 10: El gráfico muestra que, en la simulación de votos realizada, la mayoría de la población que ejerce el voto están en un rango de 26 a 40 años y de 56 años en adelante.*

### 3.2.2.3 Análisis de intención de voto por partido según la edad.

En este análisis se pretende observar la estimación de votos que tiene cada partido según el rango de edades descritas en el análisis anterior.

#### Datos utilizados:

Partidos políticos	18 a 25 años (Jóvenes)	26 a 40 años (Adulto joven)	41 a 55 años (Adultos)	56 o más (Adulto mayor)
ACCESIBILIDAD SIN EXCLUSION	1	3	3	9
ACCION CIUDADANA	121	304	246	406
ALIANZA DEMOCRATA CRISTIANA	3	3	1	7
DE LOS TRABAJADORES	0	3	3	5
FRENTE AMPLIO	6	7	9	16
INTEGRACION NACIONAL	57	99	81	176
LIBERACION NACIONAL	111	247	204	414
MOVIMIENTO LIBERTARIO	8	10	7	15
NUEVA GENERACION	6	8	13	17
RENOVACION COSTARRICENSE	2	7	3	12
REPUBLICANO SOCIAL CRISTIANO	20	71	53	125
RESTAURACION NACIONAL	146	309	268	545
UNIDAD SOCIAL CRISTIANA	78	202	167	303
VOTOS NULOS	7	11	10	23
VOTOS BLANCOS	0	2	1	3

VOTOS TOTALES	566	1286	1072	2076
---------------	-----	------	------	------

Gráfico:

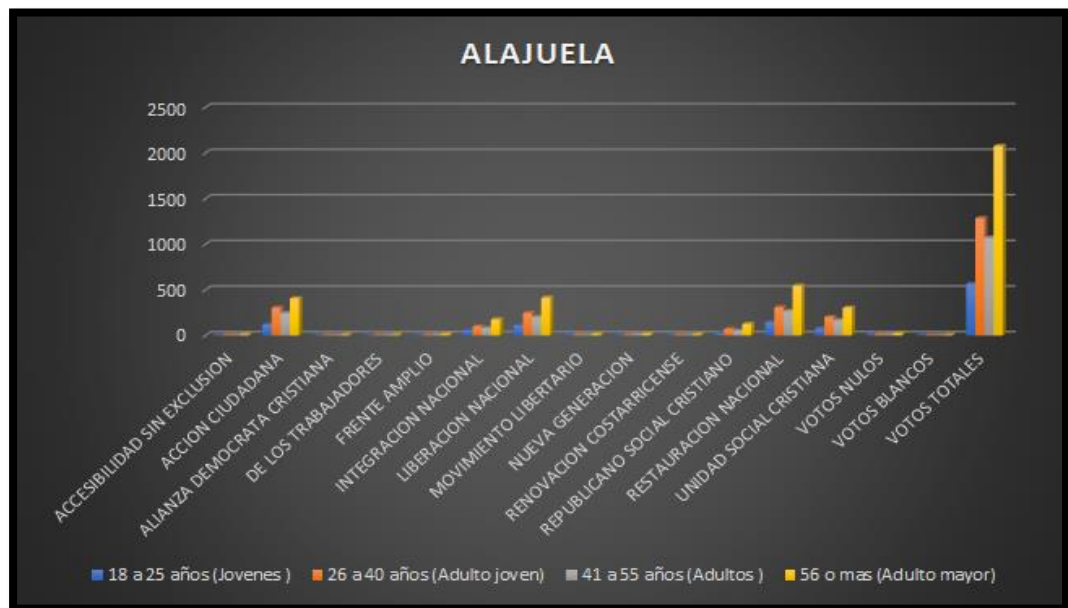


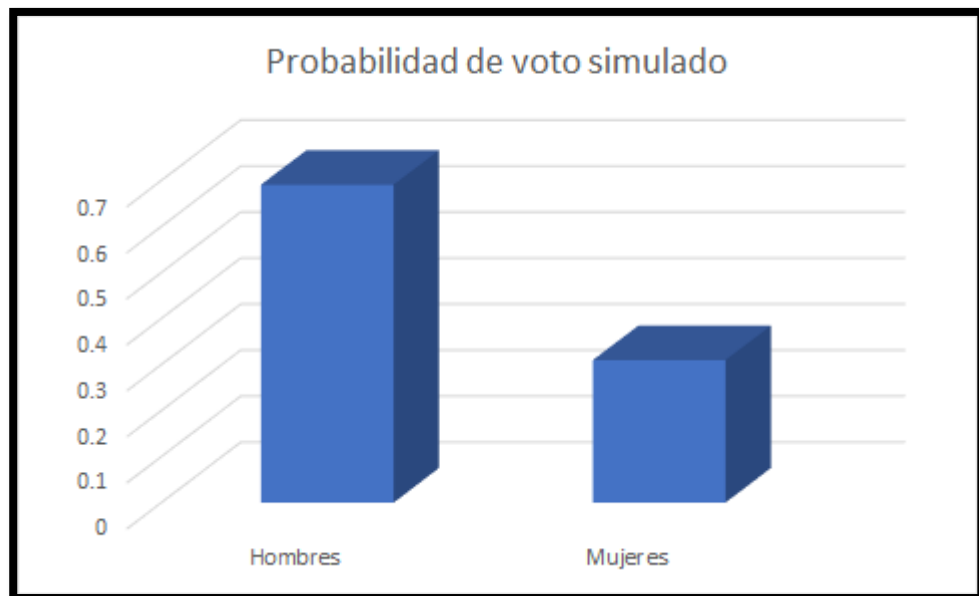
Figura 11: Gráfico que muestra una simulación de votos por partido según un rango de edad.

### 3.2.2.4 Análisis de voto por sexo.

Datos utilizados:

Género	Probabilidad de voto simulado
Hombres	0.6906
Mujeres	0.3094

Gráfico:



*Figura 12: El gráfico muestra una simulación de votos por sexo de las personas.*

### 3.2.3 Cartago.

#### 3.2.3.1 Análisis probabilidad de voto por partido.

- ❖ Probabilidad de voto real: Son las probabilidades reales basadas en los datos que el profesor brindó.
- ❖ Probabilidad de voto estimado: Para estimar esta probabilidad se implementó una función que simula 5 veces los datos y las probabilidades que dicha función retorna una vez estimados estos datos se promedian y se obtiene la probabilidad deseada. Es importante recalcar que las pruebas realizadas se hacen con una muestra de 5000 personas.

### Datos utilizados:

A continuación, se muestra una tabla con las probabilidades descritas anteriormente.

<b>Partidos políticos</b>	<b>Probabilidad de voto real</b>	<b>Probabilidad de voto simulada</b>
ACCESIBILIDAD SIN EXCLUSION	0.0004	0.0003
ACCION CIUDADANA	0.03	0.03
ALIANZA DEMOCRATA CRISTIANA	0.002	0.002
DE LOS TRABAJADORES	0.0002	0.0001
FRENTE AMPLIO	0.0009	0.0008
INTEGRACION NACIONAL	0.01	0.01
LIBERACION NACIONAL	0.02	0.02
MOVIMIENTO LIBERTARIO	0.001	0.001
NUEVA GENERACION	0.001	0.0007
RENOVACION COSTARRICENSE	0.0008	0.0005
REPUBLICANO SOCIAL CRISTIANO	0.007	0.002
RESTAURACION NACIONAL	0.02	0.02
UNIDAD SOCIAL CRISTIANA	0.02	0.02
VOTOS NULOS	0.01	0.001
VOTOS BLANCOS	0.0003	0.0001



Gráfico:

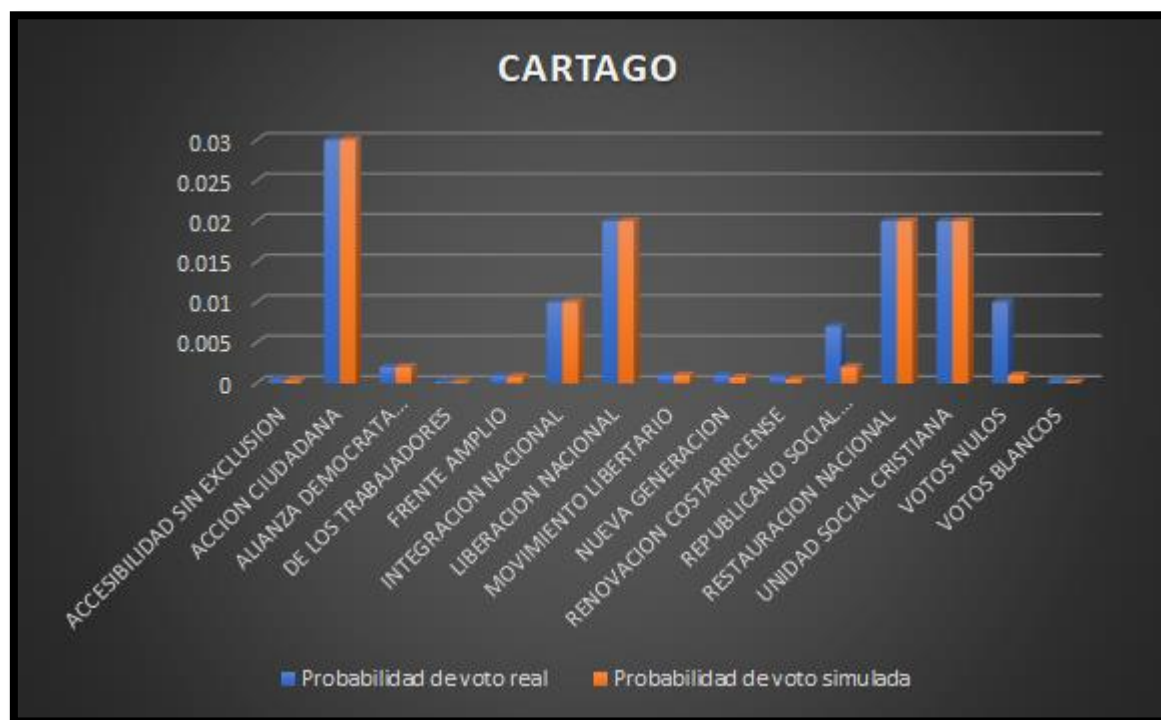


Figura 13: El gráfico mostrado hace una comparación de la probabilidad de votos reales emitidos y la probabilidad estimada por cada partido; este gráfico se obtiene según los datos mostrados anteriormente. En este gráfico se muestran diferencias un poco significativas en los votos nulos y el partido Republicano Social Cristiano.

### 3.2.3.2 Análisis de intención de voto edad.

- ❖ Probabilidad de voto estimado: Se realiza una función que estima una probabilidad de votantes que están en un rango de edades.

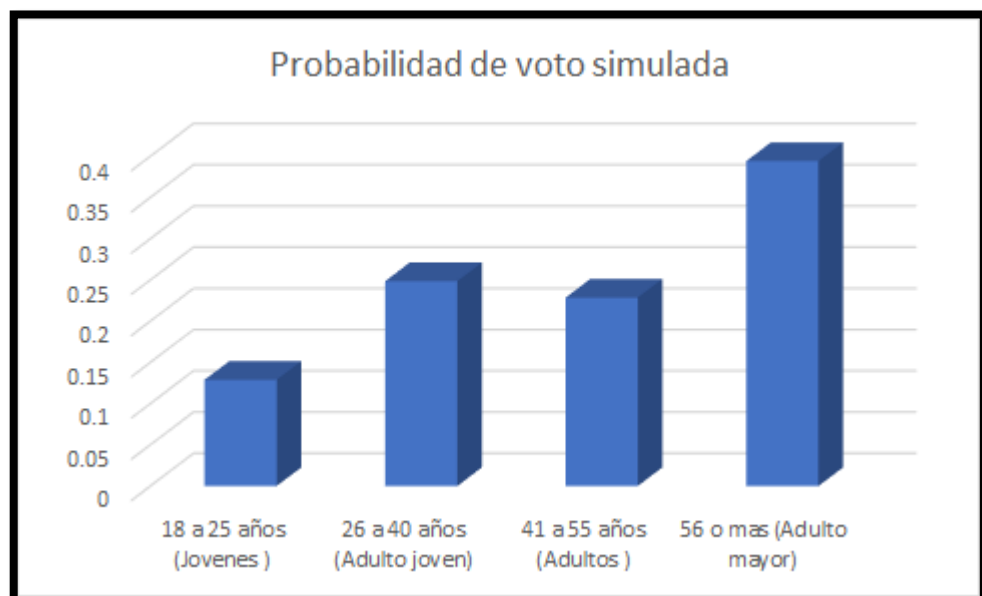
Datos utilizados:

A continuación, se muestra una tabla con las probabilidades descritas anteriormente.

Edades	Probabilidad de voto simulada
18 a 25 años (Jóvenes)	0.1288
26 a 40 años (Adulto joven)	0.2484

41 a 55 años (Adultos)	0.2286
56 o más (Adulto mayor)	0.3942

**Gráfico:**



*Figura 14: El gráfico muestra que, en la simulación de votos realizada, la mayoría de la población que ejerce el voto están en un rango de 26 a 40 años y de 56 años en adelante.*

### **3.2.3.3 Análisis de intención de voto por partido según la edad.**

En este análisis se pretende observar la estimación de votos que tiene cada partido según el rango de edades descritas en el análisis anterior.

**Datos utilizados:**

<b>Partidos políticos</b>	<b>18 a 25 años (Jóvenes)</b>	<b>26 a 40 años (Adulto joven)</b>	<b>41 a 55 años (Adultos)</b>	<b>56 o más (Adulto mayor)</b>
ACCESIBILIDAD SIN EXCLUSION	4	3	1	10
ACCION CIUDADANA	166	317	281	515
ALIANZA DEMOCRATA CRISTIANA	11	26	21	34
DE LOS TRABAJADORES	1	0	1	4
FRENTE AMPLIO	4	14	7	14
INTEGRACION NACIONAL	67	136	131	192
LIBERACION NACIONAL	138	248	225	423
MOVIMIENTO LIBERTARIO	5	13	9	16
NUEVA GENERACION	6	11	9	21
RENOVACION COSTARRICENSE	3	12	6	14
REPUBLICANO SOCIAL CRISTIANO	32	83	75	123
RESTAURACION NACIONAL	95	193	166	277
UNIDAD SOCIAL CRISTIANA	95	176	169	319
VOTOS NULOS	13	17	10	26
VOTOS BLANCOS	2	1	2	4
VOTOS TOTALES	642	1250	1133	1995

**Gráfico:**

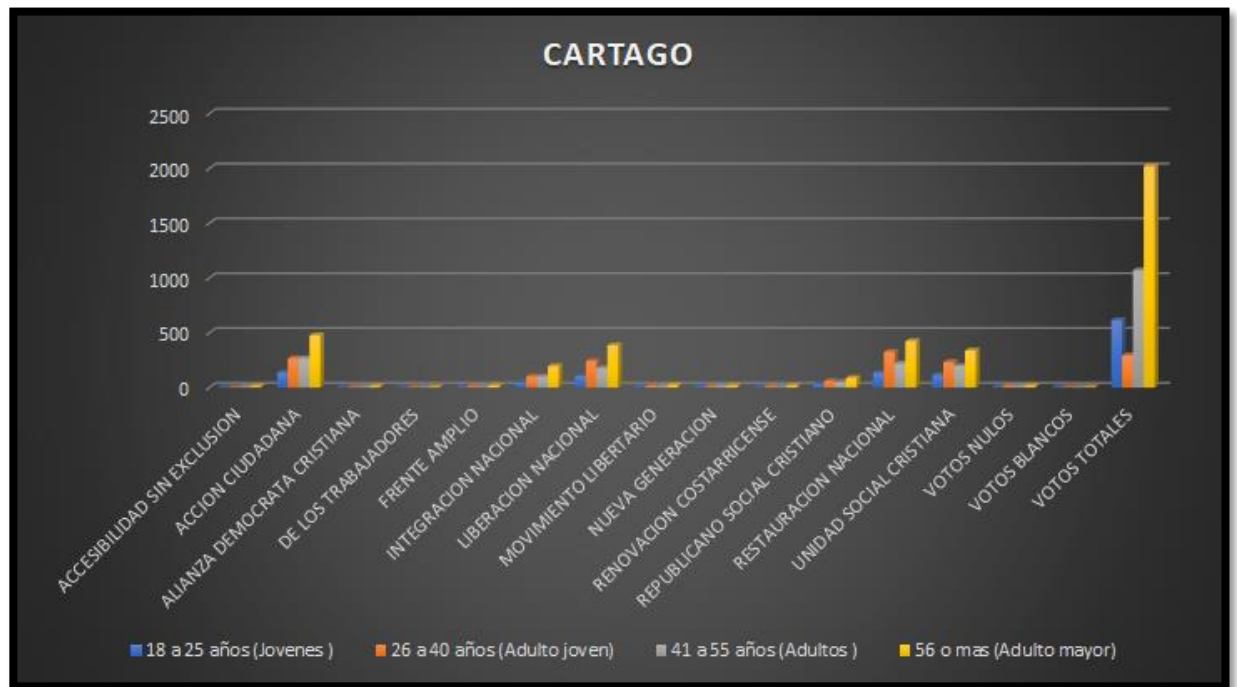


Figura 15: El gráfico muestra una simulación de votos por partido según un rango de edad.

Genero	Probabilidad de voto simulado
Hombres	0.711
Mujeres	0.289

#### 3.2.2.4 Análisis de voto por sexo.

Datos utilizados:

Gráfico:

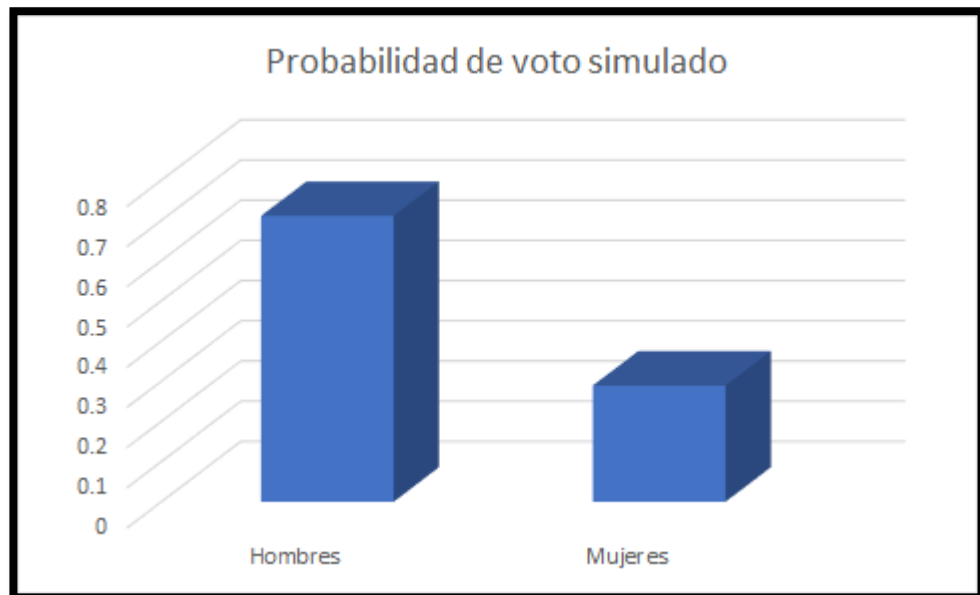


Figura 16: El gráfico muestra una simulación de votos por sexo de las personas.

### 3.2.4 Heredia.

#### 3.2.4.1 Análisis probabilidad de voto por partido.

- ❖ Probabilidad de voto real: Son las probabilidades reales basadas en los datos que el profesor brindó.
- ❖ Probabilidad de voto estimado: Para estimar esta probabilidad se implementó una función que simula 5 veces los datos y las probabilidades que dicha función retorna una vez estimados estos datos se promedian y se obtiene la probabilidad deseada. Es importante recalcar que las pruebas realizadas se hacen con una muestra de 5000 personas.

#### Datos utilizados:

A continuación, se muestra una tabla con las probabilidades descritas anteriormente.

Partidos políticos	Probabilidad de voto real	Probabilidad de voto simulada
ACCESIBILIDAD SIN EXCLUSION	0.0002	0.0002
ACCION CIUDADANA	0.03	0.03
ALIANZA DEMOCRATA CRISTIANA	0.0003	0.0002

DE LOS TRABAJADORES	0.0001	0
FRENTE AMPLIO	0.0008	0.0007
INTEGRACION NACIONAL	0.008	0.008
LIBERACION NACIONAL	0.01	0.01
MOVIMIENTO LIBERTARIO	0.001	0.0009
NUEVA GENERACION	0.001	0.0009
RENOVACION COSTARRICENSE	0.0005	0.0006
REPUBLICANO SOCIAL CRISTIANO	0.004	0.004
RESTAURACION NACIONAL	0.02	0.02
UNIDAD SOCIAL CRISTIANA	0.02	0.02
VOTOS NULOS	0.0008	0.0009
VOTOS BLANCOS	0.0001	0

### Gráfico.

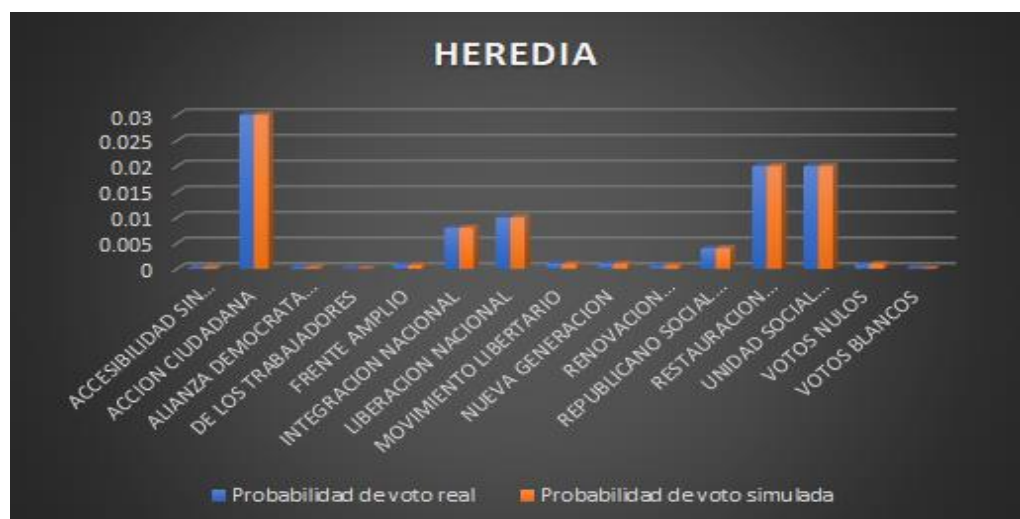


Figura 17: Gráfico que hace una comparación de la probabilidad de votos reales emitidos y la probabilidad estimada por cada partido; este gráfico se obtiene según los datos mostrados anteriormente. Además, se puede visualizar que los datos no varían mucho.

### 3.2.4.2 Análisis de intención de voto por edad.

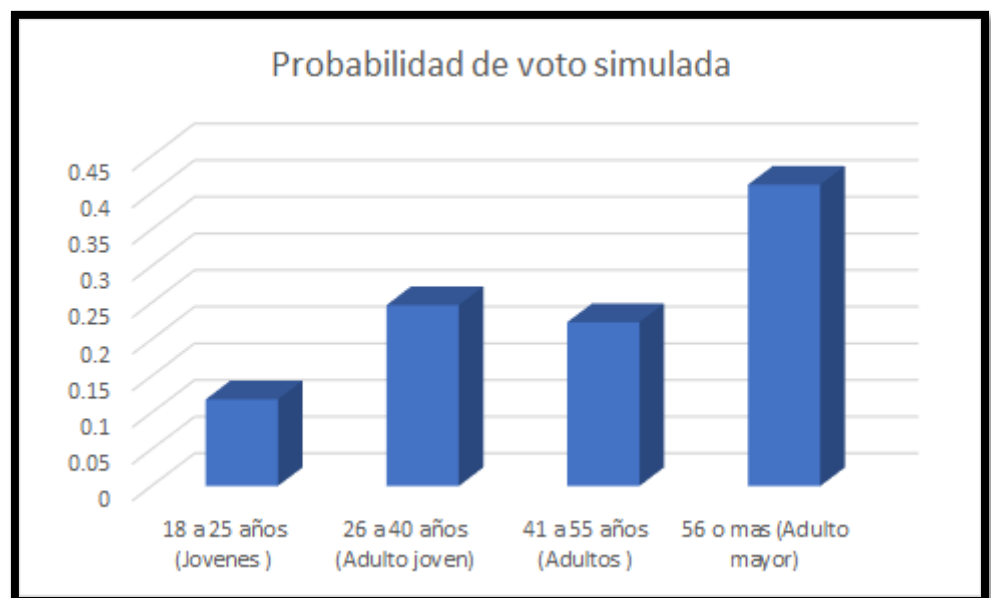
- ❖ Probabilidad de voto estimado: Se realiza una función que estima una probabilidad de votantes que están en un rango de edades.

#### Datos utilizados:

A continuación, se muestra una tabla con las probabilidades descritas anteriormente.

Edades	Probabilidad de voto simulada
18 a 25 años (Jóvenes)	0.1184
26 a 40 años (Adulto joven)	0.2468
41 a 55 años (Adultos)	0.2236
56 o más (Adulto mayor)	0.4112

#### Gráfico:



*Figura 18: El gráfico muestra que, en la simulación de votos realizada, la mayoría de la población que ejerce el voto están en un rango de 26 a 40 años y de 56 años en adelante.*

### 3.2.4.3 Análisis de intención de voto por partido según la edad.

En este análisis se pretende observar la estimación de votos que tiene cada partido según el rango de edades descritas en el análisis anterior.

#### Datos utilizados:

Partidos políticos	18 a 25 años (Jóvenes)	26 a 40 años (Adulto joven)	41 a 55 años (Adultos)	56 o más (Adulto mayor)
ACCESIBILIDAD SIN EXCLUSION	0	4	3	4
ACCION CIUDADANA	167	367	289	530
ALIANZA DEMOCRATA CRISTIANA	1	5	2	9
DE LOS TRABAJADORES	0	3	4	4
FRENTE AMPLIO	2	11	5	15
INTEGRACION NACIONAL	57	99	68	59
LIBERACION NACIONAL	121	221	186	360
MOVIMIENTO LIBERTARIO	10	11	9	28
NUEVA GENERACION	4	15	18	18
RENOVACION COSTARRICENSE	3	6	6	9
REPUBLICANO SOCIAL CRISTIANO	26	55	34	72
RESTAURACION NACIONAL	121	250	236	421
UNIDAD SOCIAL CRISTIANA	103	241	118	391
VOTOS NULOS	1	10	7	11
VOTOS BLANCOS	0	2	2	3
VOTOS TOTALES	616	1300	1050	2034



Gráfico.

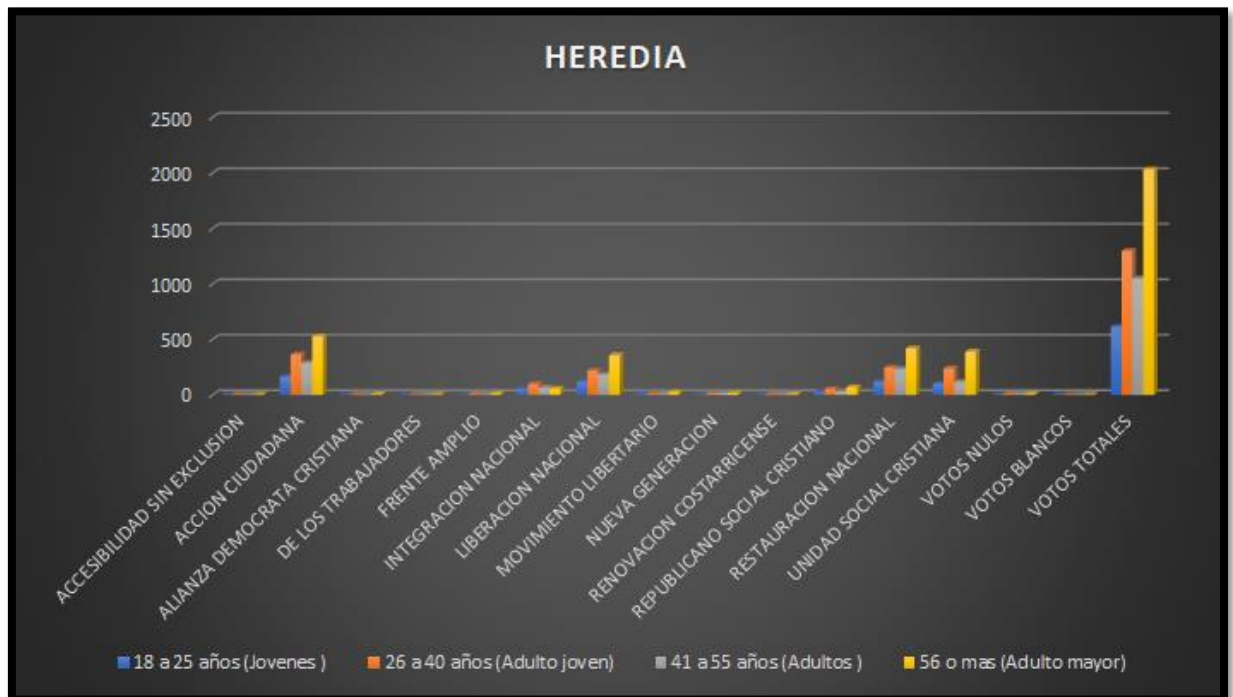


Figura 20: El gráfico muestra una simulación de votos por partido según un rango de edad.

Genero	Probabilidad de voto simulado
Hombres	0.65
Mujeres	0.35

#### 3.2.4.4 Análisis de voto por sexo.

Datos utilizados:

Gráfico:

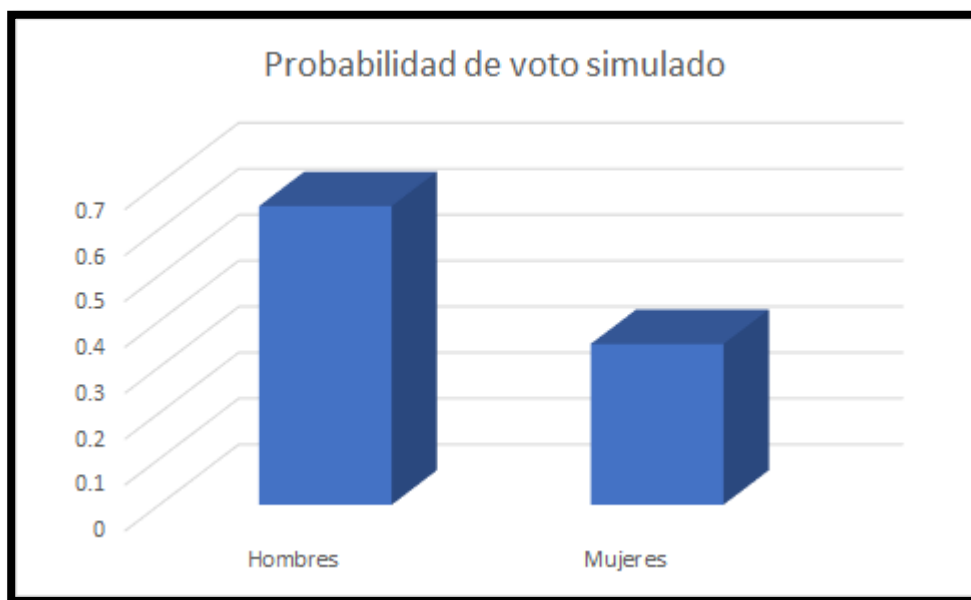


Figura 21: El gráfico muestra una simulación de votos por sexo de las personas.

### 3.2.5 Guanacaste.

#### 3.2.5.1 Análisis por partido Análisis probabilidad de voto por partido.

- ❖ Probabilidad de voto real: Son las probabilidades reales basadas en los datos que el profesor brindó.
- ❖ Probabilidad de voto estimado: Para estimar esta probabilidad se implementó una función que simula 5 veces los datos y las probabilidades que dicha función retorna una vez estimados estos datos se promedian y se obtiene la probabilidad deseada. Es importante recalcar que las pruebas realizadas se hacen con una muestra de 5000 personas.

### Datos utilizados:

A continuación, se muestra una tabla con las probabilidades descritas anteriormente.

Partidos políticos	Probabilidad de voto real	Probabilidad de voto simulada
ACCESIBILIDAD SIN EXCLUSION	0.0002	0.0001
ACCION CIUDADANA	0.001	0.001
ALIANZA DEMOCRATA CRISTIANA	0.0001	0
DE LOS TRABAJADORES	0.0001	0
FRENTE AMPLIO	0.0004	0.0004
INTEGRACION NACIONAL	0.007	0.007
LIBERACION NACIONAL	0.01	0.01
MOVIMIENTO LIBERTARIO	0.0006	0.0005
NUEVA GENERACION	0.0006	0.0002
RENOVACION COSTARRICENSE	0.0003	0.0003
REPUBLICANO SOCIAL CRISTIANO	0.003	0.003
RESTAURACION NACIONAL	0.02	0.02
UNIDAD SOCIAL CRISTIANA	0.01	0.01
VOTOS NULOS	0.01	0.0009
VOTOS BLANCOS	0.0002	0.0002

Gráfico:

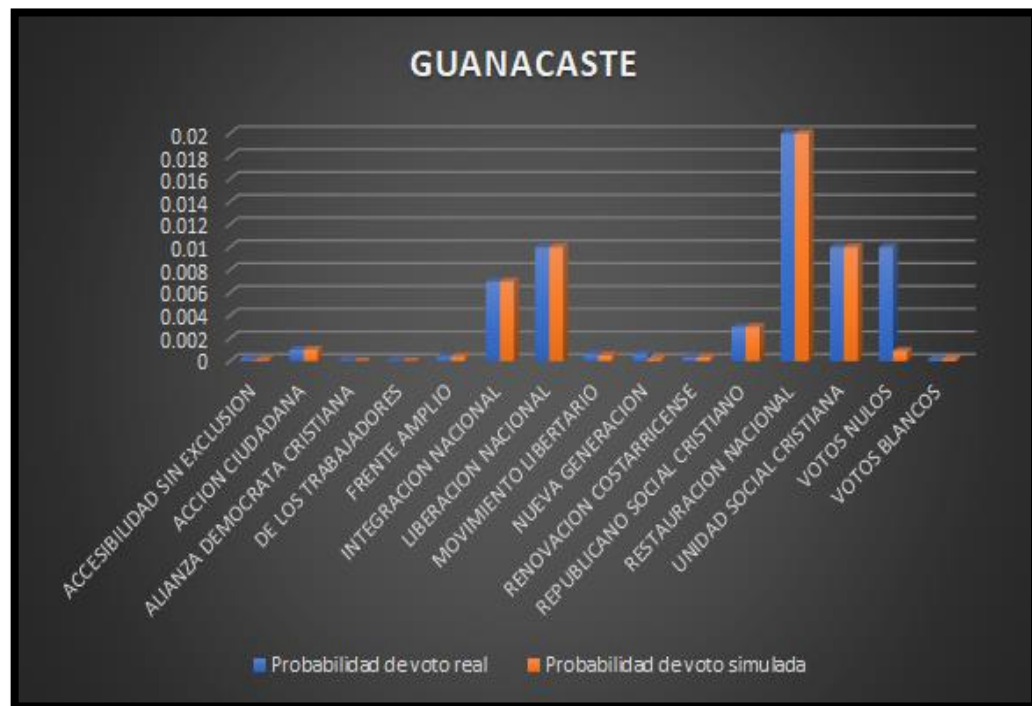


Figura 21: El gráfico mostrado hace una comparación de la probabilidad de votos reales emitidos y la probabilidad estimada por cada partido; este gráfico se obtiene según los datos mostrados anteriormente. Se puede observar que la diferencia en los votos nulos es bastante significativa.

### 3.2.5.2 Análisis de intención de voto por edad.

- ❖ Probabilidad de voto estimado: Se realiza una función que estima una probabilidad de votantes que están en un rango de edades entre 18 a 24 años y de 25 años en adelante.

### Datos utilizados:

A continuación, se muestra una tabla con las probabilidades descritas anteriormente.

Edades	Probabilidad de voto simulada
18 a 25 años (Jóvenes)	0.1234
26 a 40 años (Adulto joven)	0.2576
41 a 55 años (Adultos)	0.2134
56 o más (Adulto mayor)	0.4056

Gráfico:

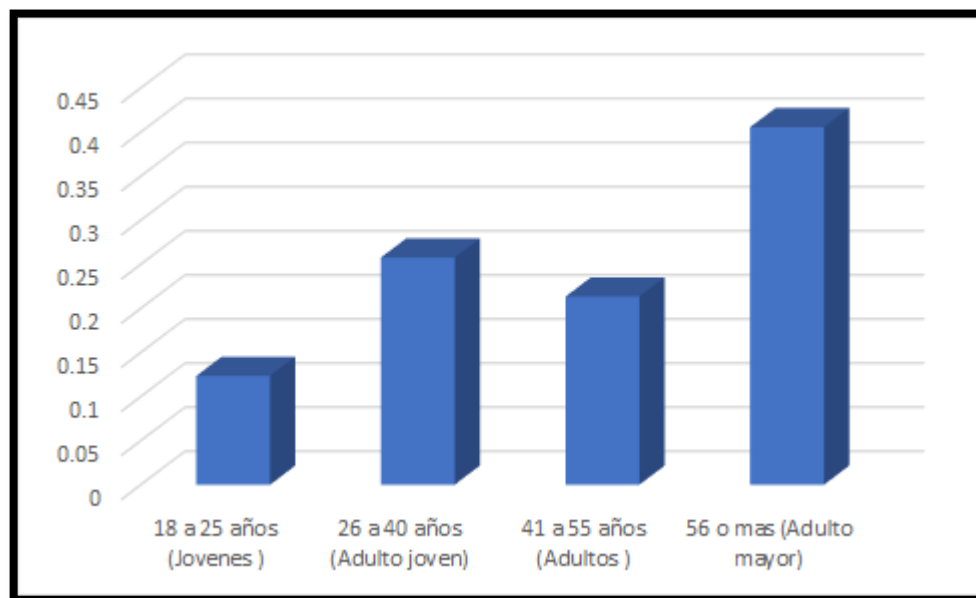


Figura 22: El gráfico muestra que, en la simulación de votos realizada, la mayoría de la población que ejerce el voto están en un rango de 26 a 40 años y de 56 años en adelante.

### 3.2.5.3 Análisis de intención de voto por partido según la edad.

En este análisis se pretende observar la estimación de votos que tiene cada partido según el rango de edades descritas en el análisis anterior.

**Datos utilizados:**

<b>Partidos políticos</b>	<b>18 a 25 años (Jóvenes)</b>	<b>26 a 40 años (Adulto joven)</b>	<b>41 a 55 años (Adultos)</b>	<b>56 o más (Adulto mayor)</b>
ACCESIBILIDAD SIN EXCLUSION	4	3	10	10
ACCION CIUDADANA	86	166	170	269
ALIANZA DEMOCRATA CRISTIANA	4	4	1	3
DE LOS TRABAJADORES	0	2	0	4
FRENTE AMPLIO	2	9	7	10
INTEGRACION NACIONAL	69	143	112	226
LIBERACION NACIONAL	136	295	253	487
MOVIMIENTO LIBERTARIO	2	16	9	20
NUEVA GENERACION	2	5	2	6
RENOVACION COSTARRICENSE	3	8	6	6
REPUBLICANO SOCIAL CRISTIANO	27	47	44	105
RESTAURACION NACIONAL	154	327	252	515
UNIDAD SOCIAL CRISTIANA	97	209	180	372
VOTOS NULOS	12	21	25	28
VOTOS BLANCOS	1	7	9	5
VOTOS TOTALES	599	1262	1073	2066

Gráfico:

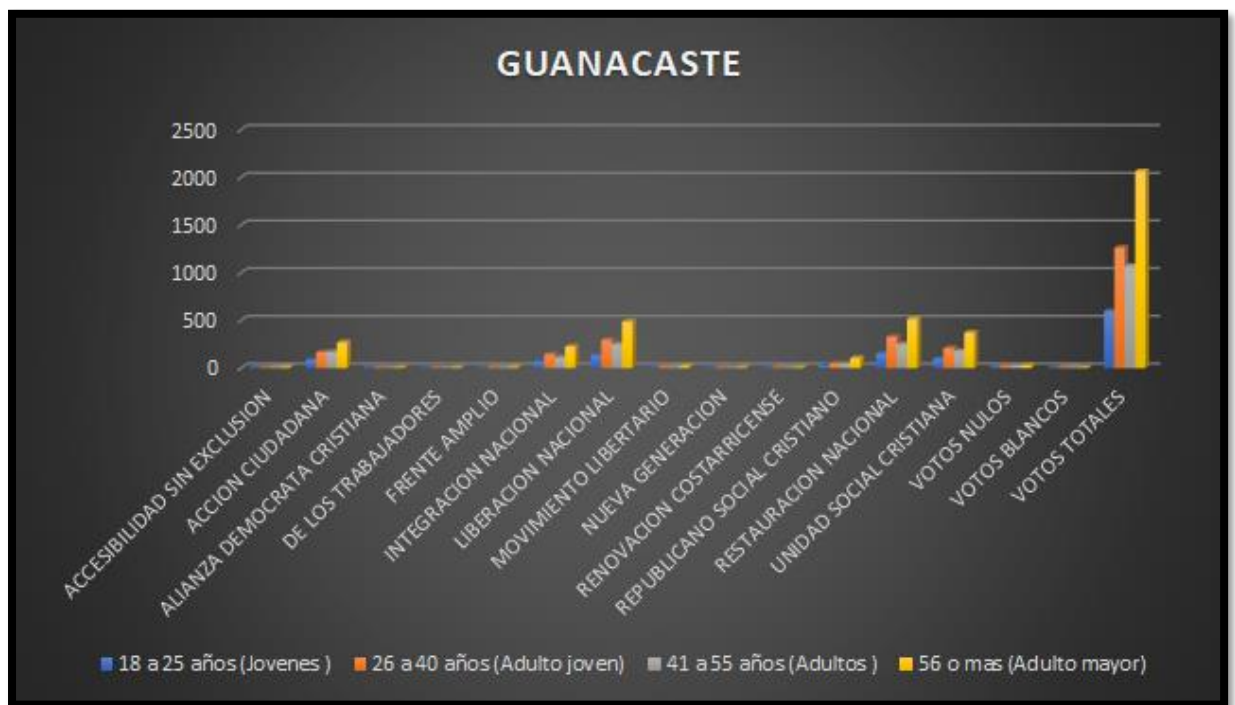


Figura 23: El gráfico muestra una simulación de votos por partido según un rango de edad.

#### 3.2.5.4 Análisis de voto por sexo.

Datos utilizados:

Género	Probabilidad de voto simulado
Hombres	0.7204
Mujeres	0.2796

Gráfico:

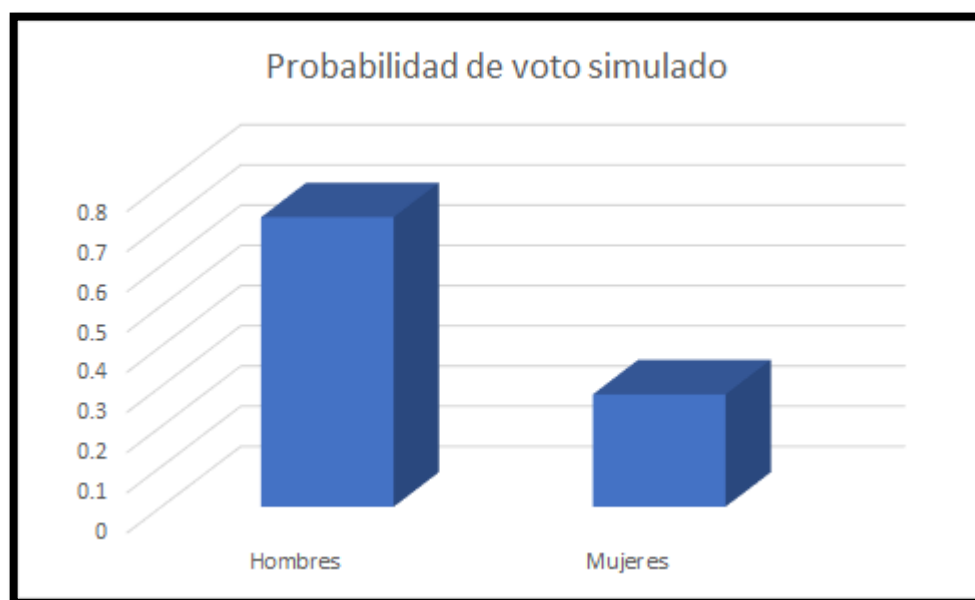


Figura 24: El gráfico muestra una simulación de votos por sexo de las personas.

### 3.2.6 Puntareñas.

#### 3.2.6.1 Análisis probabilidad de voto por partido.

- ❖ Probabilidad de voto real: Son las probabilidades reales basadas en los datos que el profesor brindó.
- ❖ Probabilidad de voto estimado: Para estimar esta probabilidad se implementó una función que simula 5 veces los datos y las probabilidades que dicha función retorna una vez estimados estos datos se promedian y se obtiene la probabilidad deseada. Es importante recalcar que las pruebas realizadas se hacen con una muestra de 5000 personas.

#### Datos utilizados:

A continuación, se muestra una tabla con las probabilidades descritas anteriormente.

Partidos políticos	Probabilidad de voto real	Probabilidad de voto simulada
ACCESIBILIDAD SIN EXCLUSION	0.0002	0
ACCION CIUDADANA	0.009	0.009
ALIANZA DEMOCRATA CRISTIANA	0.0002	0.0002



DE LOS TRABAJADORES	0.0001	0
FRENTE AMPLIO	0.0006	0.0005
INTEGRACION NACIONAL	0.009	0.009
LIBERACION NACIONAL	0.01	0.01
MOVIMIENTO LIBERTARIO	0.0008	0.0009
NUEVA GENERACION	0.0004	0.0002
RENOVACION COSTARRICENSE	0.0004	0.0004
REPUBLICANO SOCIAL CRISTIANO	0.003	0.003
RESTAURACION NACIONAL	0.03	0.03
UNIDAD SOCIAL CRISTIANA	0.1	0.009
VOTOS NULOS	0.001	0.001
VOTOS BLANCOS	0.0002	0.0002

Gráfico:

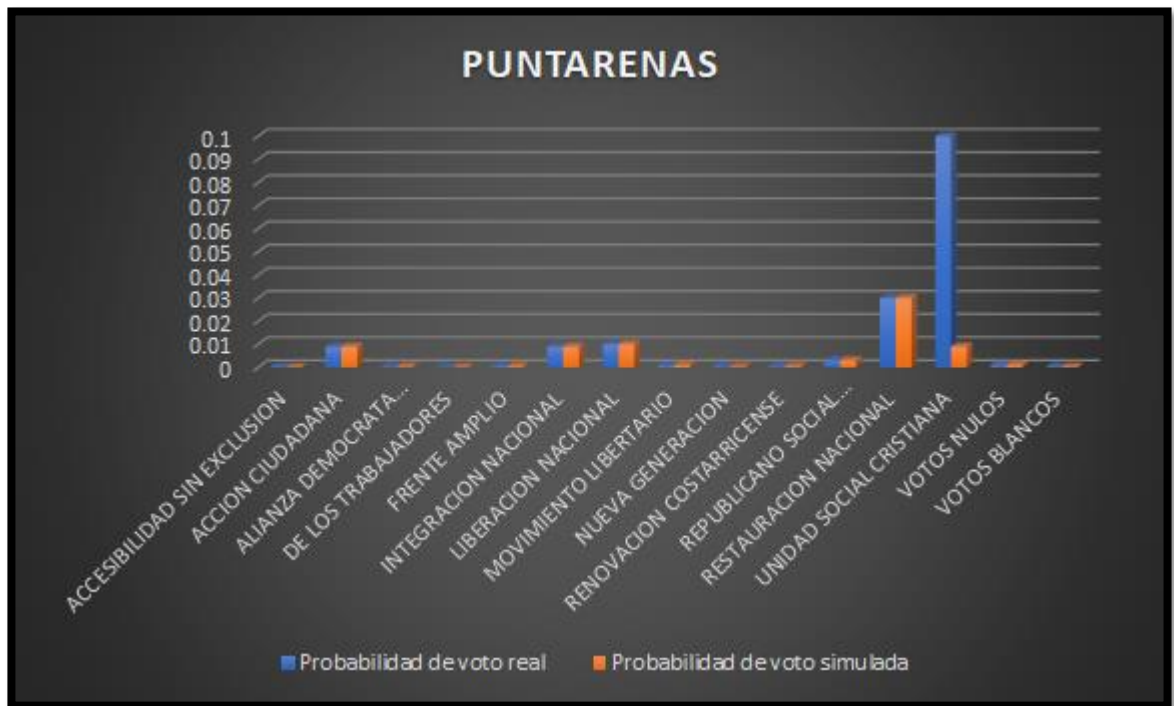


Figura 25: El gráfico mostrado hace una comparación de la probabilidad de votos reales emitidos y la probabilidad estimada por cada partido; este gráfico se obtiene según los datos mostrados anteriormente. En este gráfico es importante recalcar la diferencia que se da en el partido Unidad Social Cristiana; en cuanto a los demás partidos, se puede observar que los datos son similares.

### 3.2.6.2 Análisis de intención de voto por edad.

- ❖ Probabilidad de voto estimado: Se realiza una función que estima una probabilidad de votantes que están en un rango de edades entre 18 a 24 años y de 25 años en adelante.

#### Datos utilizados:

A continuación, se muestra una tabla con las probabilidades descritas anteriormente.

Edades	Probabilidad de voto simulada
18 a 25 años (Jóvenes)	0.118
26 a 40 años (Adulto joven)	0.2662

41 a 55 años (Adultos)	0.2142
56 o más (Adulto mayor)	0.4016

**Gráfico:**

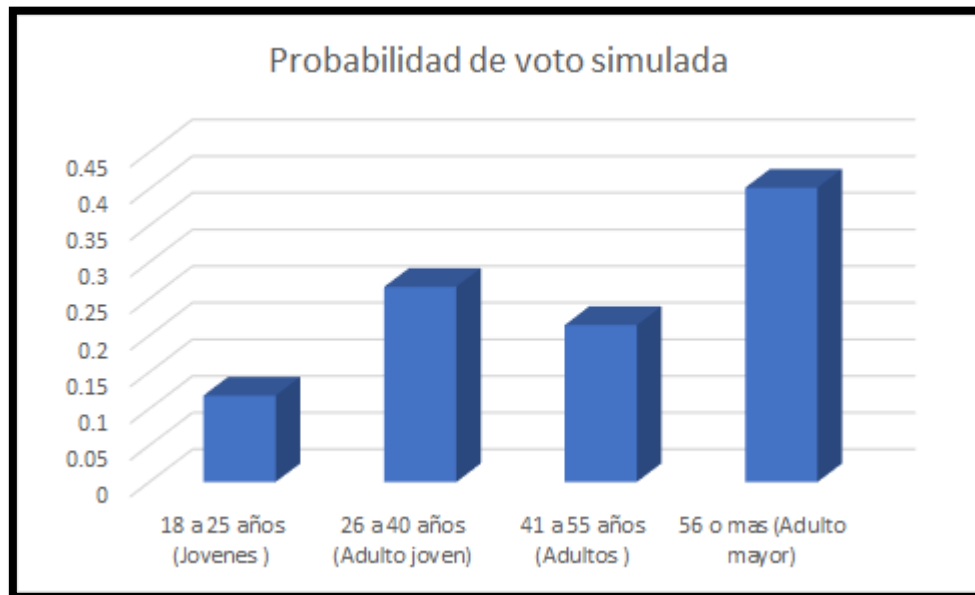


Figura 26: El gráfico muestra que, en la simulación de votos realizada, la mayoría de la población que ejerce el voto están en un rango de 26 a 40 años y de 56 años en adelante.

### 3.2.3.6 Análisis de intención de voto por partido según la edad.

En este análisis se pretende observar la estimación de votos que tiene cada partido según el rango de edades descritas en el análisis anterior.

**Datos utilizados:**

Partidos políticos	18 a 25 años (Jóvenes)	26 a 40 años (Adulto joven)	41 a 55 años (Adultos)	56 o más (Adulto mayor)
ACCESIBILIDAD SIN EXCLUSION	2	3	3	13
ACCION CIUDADANA	77	137	18	251

ALIANZA DEMOCRATA CRISTIANA	3	4	4	7
DE LOS TRABAJADORES	2	2	2	5
FRENTE AMPLIO	3	12	6	10
INTEGRACION NACIONAL	75	115	120	220
LIBERACION NACIONAL	117	211	229	353
MOVIMIENTO LIBERTARIO	11	22	11	18
NUEVA GENERACION	3	4	9	14
RENOVACION COSTARRICENSE	3	10	6	10
REPUBLICANO SOCIAL CRISTIANO	50	64	53	101
RESTAURACION NACIONAL	212	434	381	739
UNIDAD SOCIAL CRISTIANA	81	174	163	237
VOTOS NULOS	9	22	15	24
VOTOS BLANCOS	3	3	3	7
VOTOS TOTALES	651	1217	1123	2009

Gráfico:

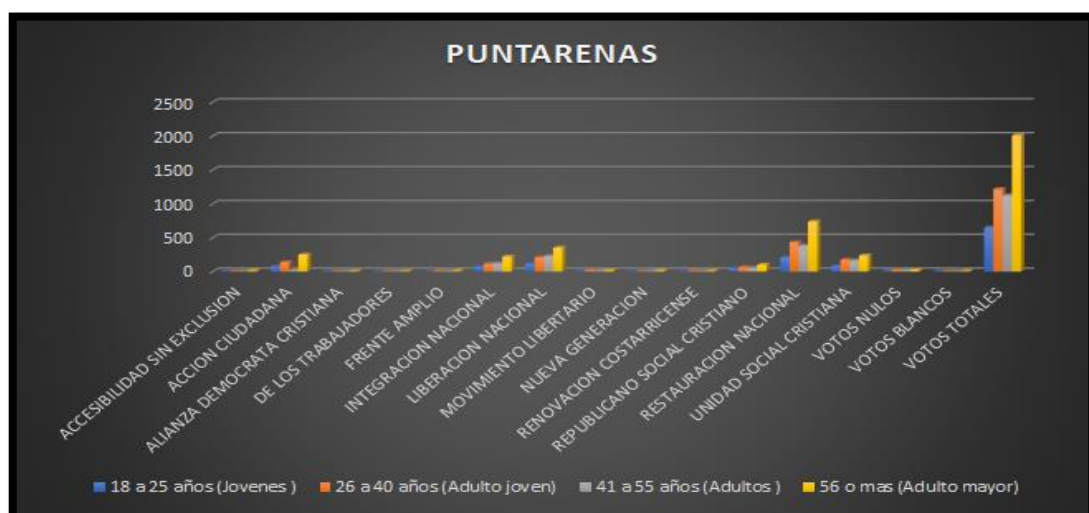


Figura 27: El gráfico muestra una simulación de votos por partido según un rango de edad.

#### 3.2.6.4 Análisis de voto por sexo.

Datos utilizados:

Género	Probabilidad de voto simulado
Hombres	0.7518
Mujeres	0.2482

Gráfico:

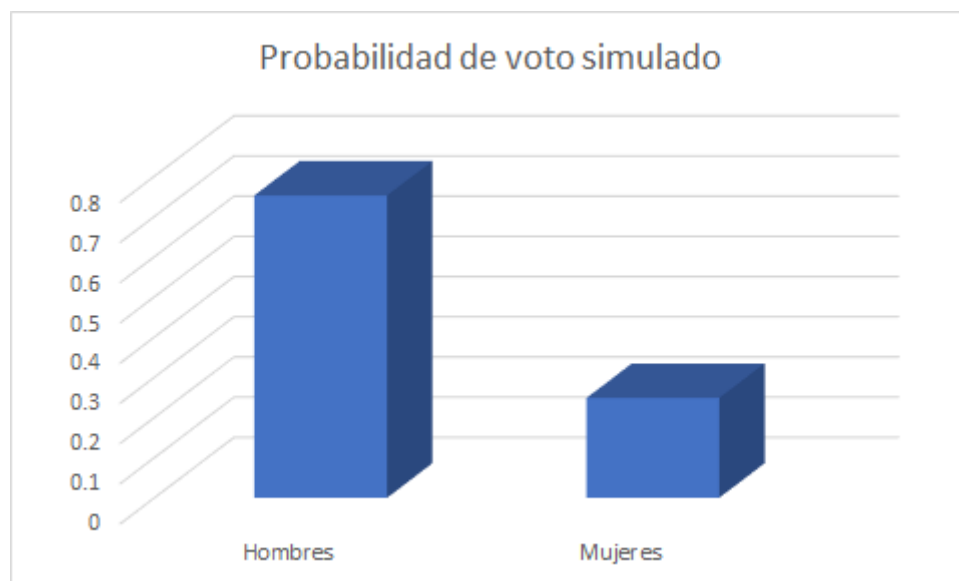


Figura 28: El gráfico muestra una simulación de votos por sexo de las personas.

#### 3.2.7 Limón.

##### 3.2.7.1 Análisis probabilidad de voto por partido.

- ❖ Probabilidad de voto real: Son las probabilidades reales basadas en los datos que el profesor brindó.
- ❖ Probabilidad de voto estimado: Para estimar esta probabilidad se implementó una función que simula 5 veces los datos y las probabilidades

que dicha función retorna una vez estimados estos datos se promedian y se obtiene la probabilidad deseada. Es importante recalcar que las pruebas realizadas se hacen con una muestra de 5000 personas.

#### Datos utilizados:

A continuación, se muestra una tabla con las probabilidades descritas anteriormente.

Partidos políticos	Probabilidad de voto real	Probabilidad de voto simulada
ACCESIBILIDAD SIN EXCLUSION	0.0005	0.0005
ACCION CIUDADANA	0.007	0.007
ALIANZA DEMOCRATA CRISTIANA	0.0004	0.0004
DE LOS TRABAJADORES	0.0002	0.0002
FRENTE AMPLIO	0.0007	0.0006
INTEGRACION NACIONAL	0.0007	0.007
LIBERACION NACIONAL	0.01	0.01
MOVIMIENTO LIBERTARIO	0.0009	0.001
NUEVA GENERACION	0.0003	0.0001
RENOVACION COSTARRICENSE	0.0005	0.0004
REPUBLICANO SOCIAL CRISTIANO	0.003	0.003
RESTAURACION NACIONAL	0.03	0.03
UNIDAD SOCIAL CRISTIANA	0.006	0.005
VOTOS NULOS	0.001	0.001
VOTOS BLANCOS	0.0003	0.0003

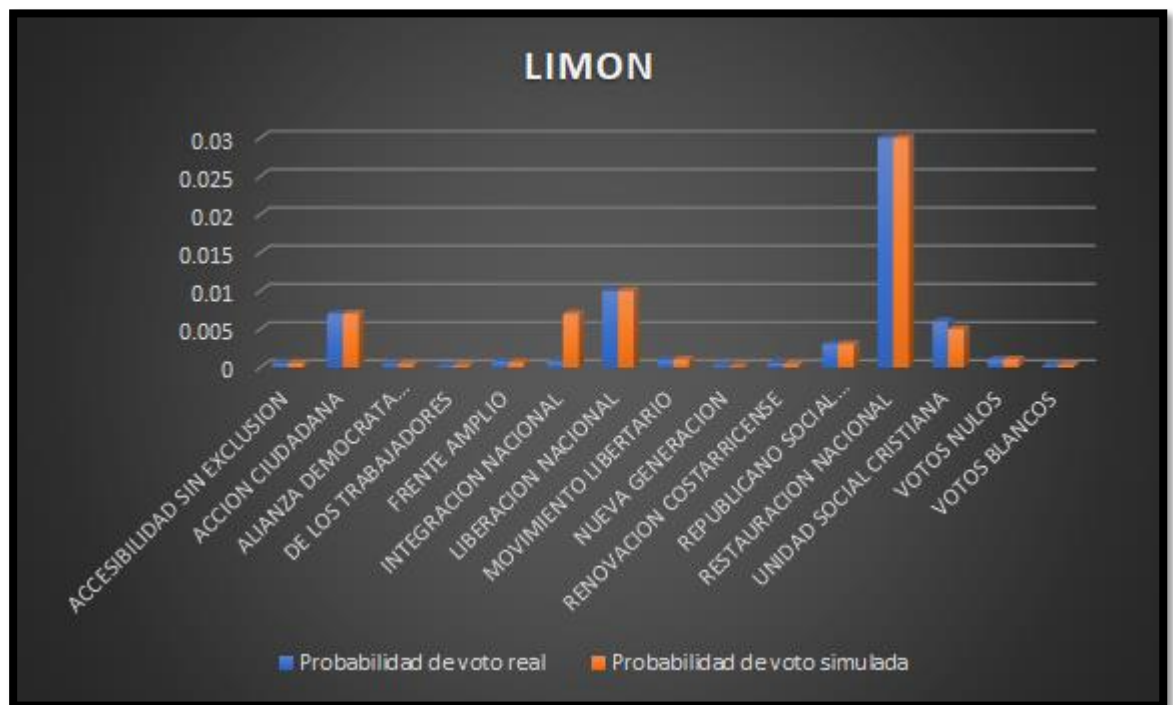


Gráfico:

### 3.2.7.2 Análisis de intención de voto por edad.

- ❖ Probabilidad de voto estimado: Se realiza una función que estima una probabilidad de votantes que están en un rango de edades entre 18 a 24 años y de 25 años en adelante.

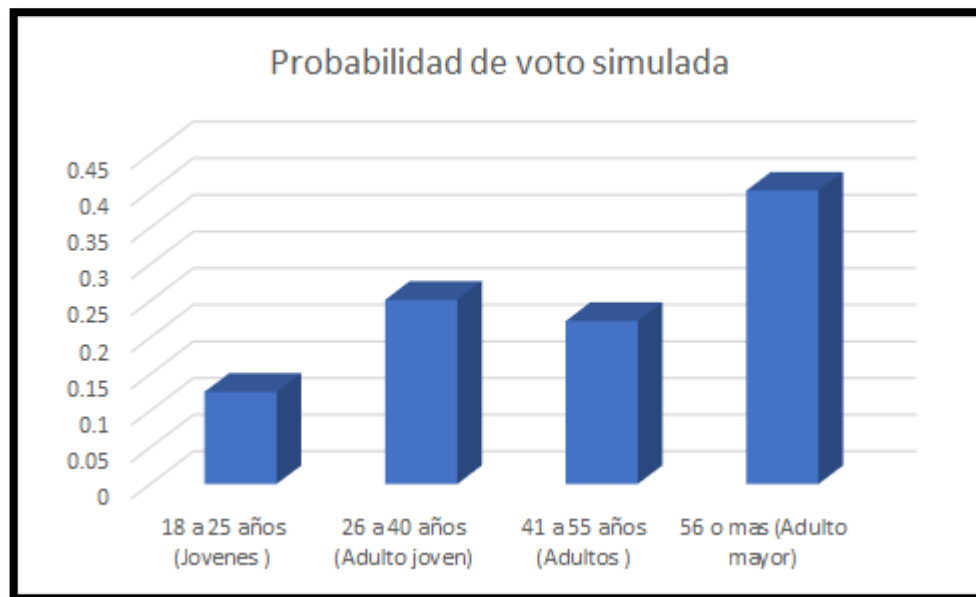
#### Datos utilizados

Edades	Probabilidad de voto simulada
18 a 25 años (Jóvenes)	0.1258
26 a 40 años (Adulto joven)	0.2514
41 a 55 años (Adultos)	0.2224

Figura 29: El gráfico mostrado hace una comparación de la probabilidad de votos reales emitidos y la probabilidad estimada por cada partido; este gráfico se obtiene según los datos mostrados anteriormente. Además, se puede visualizar que los datos no varían mucho y las diferencias que se dan no son tan significantes; es decir, no se alteran demasiado.

56 o más (Adulto mayor)	0.4004
-------------------------	--------

**Gráfico:**



*Figura 30: El gráfico muestra que, en la simulación de votos realizada, la mayoría de la población que ejerce el voto están en un rango de 26 a 40 años y de 56 años en adelante.*

### 3.2.7.3 Análisis de intención de voto por partido según la edad.

En este análisis se pretende observar la estimación de votos que tiene cada partido según el rango de edades descritas en el análisis anterior.

#### Datos utilizados:

Partidos políticos	18 a 25 años (Jóvenes)	26 a 40 años (Adulto joven)	41 a 55 años (Adultos)	56 o más (Adulto mayor)
ACCESIBILIDAD SIN EXCLUSION	3	10	9	17
ACCION CIUDADANA	61	125	115	245
ALIANZA DEMOCRATA CRISTIANA	4	9	5	14
DE LOS TRABAJADORES	1	4	6	5



FRENTE AMPLIO	8	16	10	22
INTEGRACION NACIONAL	62	140	98	191
LIBERACION NACIONAL	101	212	200	378
MOVIMIENTO LIBERTARIO	5	9	10	25
NUEVA GENERACION	1	4	5	4
RENOVACION COSTARRICENSE	5	8	8	18
REPUBLICANO SOCIAL CRISTIANO	28	57	52	88
RESTAURACION NACIONAL	257	530	452	826
UNIDAD SOCIAL CRISTIANA	61	100	106	151
VOTOS NULOS	10	16	25	48
VOTOS BLANCOS	3	6	1	10
VOTOS TOTALES	610	1246	1102	2042

**Gráfico:**

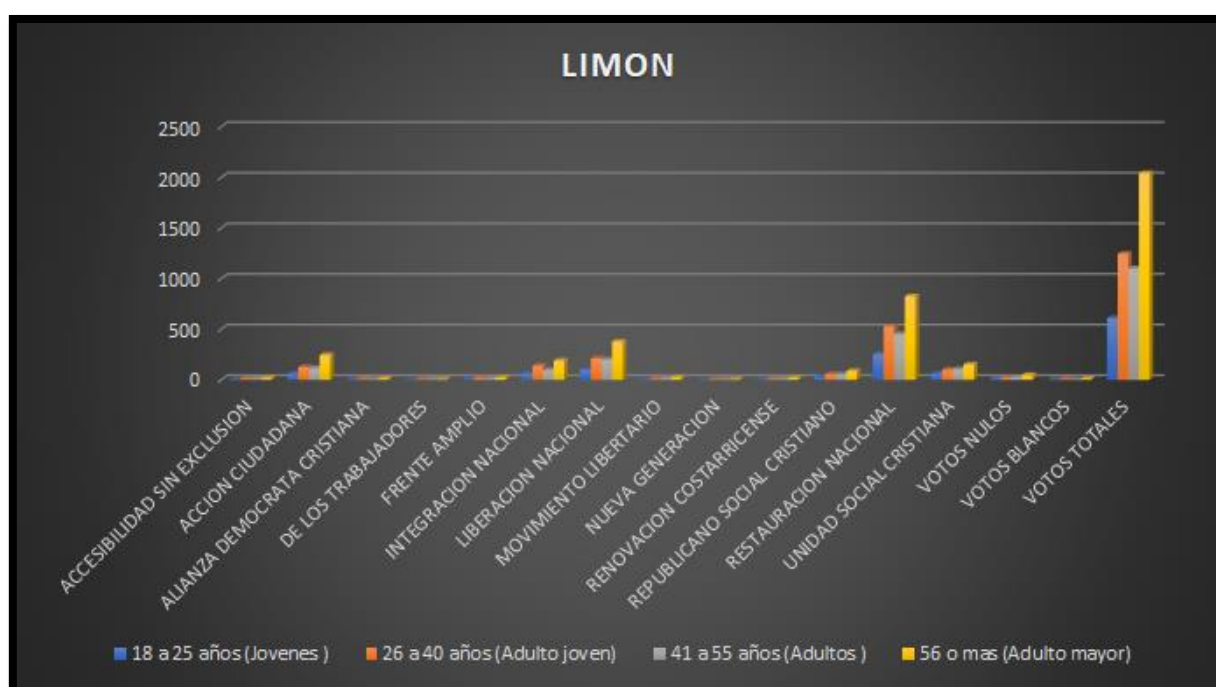


Figura 31: Gráfico que muestra una simulación de votos por partido según un rango de edad.

#### 3.2.7.4 Análisis de voto por sexo.

Datos utilizados:

Género	Probabilidad de voto simulado
Hombres	0.7438
Mujeres	0.2562

Gráfico:

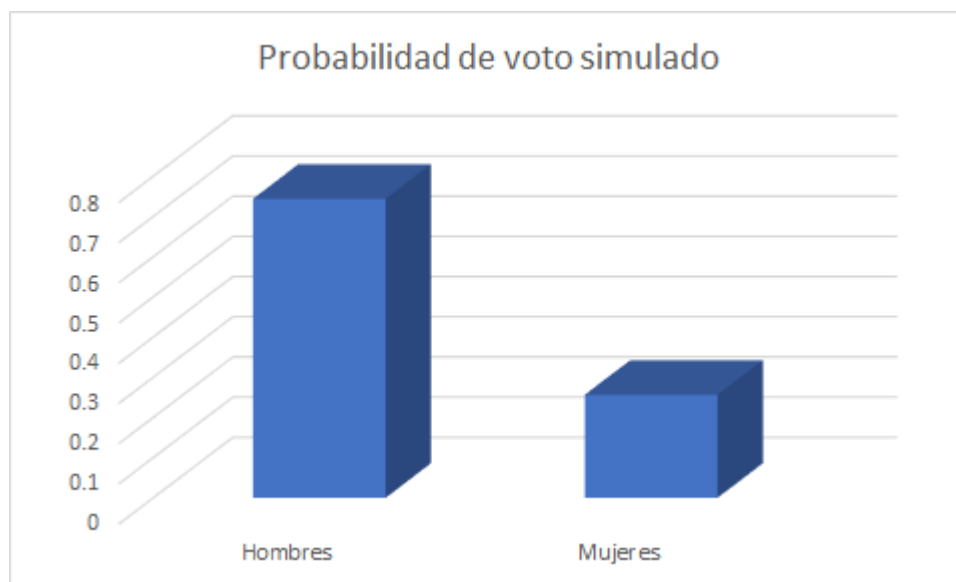


Figura 32: El gráfico muestra una simulación de votos por sexo de las personas.

#### 4. Conclusión.

---

Una vez concluido este proyecto corto se puede ver que las muestras generadas por cada una de las provincias son muy similares a los datos reales brindados por el profesor. Para verificar que los datos simulados se acercarán a la realidad, se ejecutó el código aproximadamente 5 veces y se calculó la probabilidad de votos en diferentes categorías (partido, nivel de educación, edad, sexo) una vez calculadas dichas probabilidades, se compararon con los datos reales mencionados anteriormente. Sin embargo, si se observan los gráficos del análisis de probabilidad de voto por partido, se puede ver que en algunas provincias si se dan diferencias significativas en los datos, ya sea que la probabilidad real es más alta que la simulada o viceversa.

Por otra parte, según el análisis de intención de voto por edad, se llegó a la conclusión de que, en la simulación realizada, la mayoría de los votos ejercidos son por personas que están en un rango de edad entre 26 a 40 años y de 56 años en adelante. Además, se puede ver que la mayoría de las provincias los votantes son principalmente hombres.