Jockey Nutrition: AI-Driven Insights into Food Choices and Bodyweight Management

Academic Supervisor: Teo Susnjak

Project owner / subject experts: Drs. Kylie Legg and Yin Chin

Background:

Jockeys face intense pressure to maintain low bodyweight, often without formal education in nutrition. This project explores how apprentice jockeys manage their diet, sleep, and workload over time, and whether educational interventions improve their nutritional habits and health outcomes. The data includes weekly food diaries, work diaries (horses ridden), and sleep diaries collected every 6 weeks over 3 years in a written form.

Aim:

To analyse and visualise the eating behaviours and food choices of apprentice jockeys using AI and LLM technologies. The project will identify patterns that support or hinder sustainable performance and implement Agentic AI approaches to extracting data from written text and to orchestrate machine learning models.

Objectives:

1. Classify food types (e.g., convenience foods vs. nutritious foods, food groups, colour diversity on plate).

2. Identify eating patterns (e.g., pre/post-race day, feast/famine cycles).

3. Implement LLM and Agentic AI based approaches as appropriate for automating the extraction of written text from diaries and storing them into a structured dataset

4. Implement machine learning models to extract insights from the datasets on diaries

5. Implement a Streamlit dashboard to visualise the insights

Additional sub goals:

1. Estimate nutritional intake using AI models.
2. Correlate food choices with workload.
3. Evaluate changes over time in diet quality and balance.
4. Visualize trends and generate insights for sustainable weight management.

Data Sources:

- Food Diaries: Weekly logs of meals, snacks, and beverages.

- Work Diaries: Number of horses ridden at trackwork and races.

Deliverables:

1. **Data Pipeline**:

   - LLM-based / Agentic AI-based extraction and classification of diary entries from Optical Character Recognition (OCR) through to text correction.

2. **Predictive models and a report:**

   - All project code, including code for training models uploaded to Github

   - Jupyter notebooks

   - Final report

3. **Interactive Dashboard**:

   - Visualizations of food group intake, balance, and trends.

Additional Insights to Explore once ML code established :

- Meal timing and frequency (e.g., intermittent fasting patterns).

- Cluster analysis of jockeys by behaviour patterns.

- Anomaly detection for unhealthy eating habits.

Example data for OCR extraction:
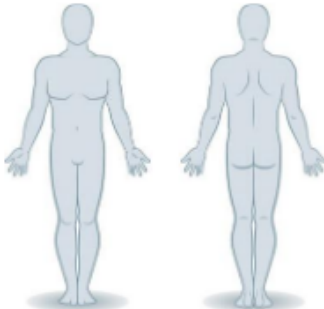
Name: _____

Date: _____

My Riding Diary

Monday	Horses ridden	Falls
Trackwork		
Gallops		
Jumpouts		
Trials		
Races		
Other eg. gym/run		

Tuesday	Horses ridden	Falls
Trackwork		
Gallops		
Jumpouts		
Trials		
Races		
Other eg. gym/run		

Wednesday	Horses ridden	Falls
Trackwork		
Gallops		
Jumpouts		
Trials		
Races		
Other eg. gym/run		

Thursday	Horses ridden	Falls
Trackwork		
Gallops		
Jumpouts		
Trials		
Races		
Other eg. gym/run		

Friday	Horses ridden	Falls
Trackwork		
Gallops		
Jumpouts		
Trials		
Races		
Other eg. gym/run		

Saturday	Horses ridden	Falls
Trackwork		
Gallops		
Jumpouts		
Trials		
Races		
Other eg. gym/run		

Sunday	Horses ridden	Falls
Trackwork		
Gallops		
Jumpouts		
Trials		
Races		
Other eg. gym/run		

Circle areas affected in falls:

Fitness focus: _____

Name: _____

Date:_____

My Food Diary

Monday	
Breakfast	
Snack	
Lunch	
Snack	
Dinner	
Snack	

Tuesday	
Breakfast	
Snack	
Lunch	
Snack	
Dinner	
Snack	

Wednesday	
Breakfast	
Snack	
Lunch	
Snack	
Dinner	
Snack	

Thursday	
Breakfast	
Snack	
Lunch	
Snack	
Dinner	
Snack	

Friday	
Breakfast	
Snack	
Lunch	
Snack	
Dinner	
Snack	

Saturday	
Breakfast	
Snack	
Lunch	
Snack	
Dinner	
Snack	

Sunday	
Breakfast	
Snack	
Lunch	
Snack	
Dinner	
Snack	

Notes:

CDC

Sleep Diary: Morning

Monday	Tuesday	Wednesday	Thursday	Friday	Saturday	Sunday
Last night, I went to bed at:						
AM/PM	AM/PM	AM/PM	AM/PM	AM/PM	AM/PM	AM/PM
This morning, I woke up at:						
AM/PM	AM/PM	AM/PM	AM/PM	AM/PM	AM/PM	AM/PM
Last night, it took me about ___ mins to fall asleep:						
mins	mins	mins	mins	mins	mins	mins
I felt the quality of my sleep was: Eg. Very good (VG), Good (G), Average (A), Bad, (B), Very Bad (VB)						
VG / G / A / B / VB	VG / G / A / B / VB	VG / G / A / B / VB	VG / G / A / B / VB	VG / G / A / B / VB	VG / G / A / B / VB	VG / G / A / B / VB
This morning I feel: Eg. Refreshed (Ref), Tired (Tir), Groggy (Grog), Alert (Al)						
Ref / Tir / Grog / Al	Ref / Tir / Grog / Al	Ref / Tir / Grog / Al	Ref / Tir / Grog / Al	Ref / Tir / Grog / Al	Ref / Tir / Grog / Al	Ref / Tir / Grog / Al
My sleep was made more difficult by: Eg. Temperature, noise, dreams, thoughts, not feeling tired, discomfort						
During the night, I woke up _____ times:						
times	times	times	times	times	times	times

Sleep Diary: Night

Monday	Tuesday	Wednesday	Thursday	Friday	Saturday	Sunday
I took a nap:						
Yes / No	Yes / No	Yes / No	Yes / No	Yes / No	Yes / No	Yes / No
I had caffeine:						
# of drinks ☐ Morning ☐ Afternoon ☐ Evening	# of drinks ☐ Morning ☐ Afternoon ☐ Evening	# of drinks ☐ Morning ☐ Afternoon ☐ Evening	# of drinks ☐ Morning ☐ Afternoon ☐ Evening	# of drinks ☐ Morning ☐ Afternoon ☐ Evening	# of drinks ☐ Morning ☐ Afternoon ☐ Evening	# of drinks ☐ Morning ☐ Afternoon ☐ Evening
Today I exercised for _____ minutes:						
Mins/hrs	Mins/hrs	Mins/hrs	Mins/hrs	Mins/hrs	Mins/hrs	Mins/hrs
Medications or drugs I used today:						
Throughout the day, I felt drowsy:						
☐ Never ☐ Sometimes ☐ Very often	☐ Never ☐ Sometimes ☐ Very often	☐ Never ☐ Sometimes ☐ Very often	☐ Never ☐ Sometimes ☐ Very often	☐ Never ☐ Sometimes ☐ Very often	☐ Never ☐ Sometimes ☐ Very often	☐ Never ☐ Sometimes ☐ Very often
Overall, my mood today was: Eg. Positive (Pos), Negative (Neg), Neutral (Neu)						
Pos / Neg / Neu	Pos / Neg / Neu	Pos / Neg / Neu	Pos / Neg / Neu	Pos / Neg / Neu	Pos / Neg / Neu	Pos / Neg / Neu
In the hour before bed, my activities included: Eg. Reading, computer, TV, showering, phone, eating, spending time with partner						

Example data for Machine Learning:

		Name	Jockey 1	Jockey 1	Jockey 1	Jockey 1
Apprentce Jockey Data 2023 Food Diary		**Test Number**	1	2	3	4
		Test Date	20/03/2023	1/05/2023	12/06/2023	17/07/2023
		Region	CD	CD	CD	CD
Monday		**Breakfast**	pie	mince on toast		muesli bar
		Snack1				
		Lunch	Chicken, rice, salad	Wraps with salad, cheese, egg	Hastings sandwhich	sushi
		Snack2				
		Dinner			Takeaway burger and chips	stirfry
		Snack3				
Tuesday		**Breakfast**		cous cous with veggies	Toast with peanut butter	muesli bar
		Snack1				
		Lunch	Sandwich	pear		pizza reheat
		Snack2			A couple of lollies	
		Dinner	Steak, salad	meat and veggies	Steak, roast, veg, salad	meat and veges
		Snack3				
Wednesday		**Breakfast**	toast	sushi	banana	apple
		Snack1				sausage roll
		Lunch	yoghurt	chicken wrap		
		Snack2				gingernut
		Dinner	salad, egg, potato	rice with veges, meat		meat, veges, bread
		Snack3				
Thursday		**Breakfast**	egg on toast		muesli bar	banana
		Snack1		crackers		
		Lunch	bun	chicken	salad	
		Snack2			chocolate	m&M's
		Dinner	fish, salad		chicken wrap	burger, chips
		Snack3				
Friday		**Breakfast**	fish		banana	banana
		Snack1		meat sticks		
		Lunch				
		Snack2				
		Dinner				
		Snack3				
Saturday		**Breakfast**		yoghurt	yoghurt	muesli bar
		Snack1				

		Col1	Col2	Col3	Col4
	Lunch	noodles, chicken			salad
	Snack2			chocolate	grapes, chocolate
	Dinner	McD's	pulled pork buns	burger and salad	pizza
	Snack3				
Sunday	**Breakfast**	egg, toast, potato	yoghurt	Toast with egg	toast, sausage, spaghetti
	Snack1				muesli bar
	Lunch		chicken wrap	leftover fried rice	
	Snack2	crackers	chocolate		apple
	Dinner	chicken wrap	pizza	chicken, potatoes, veges, apple pie, icecream	
	Snack3				